

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Clinical Neurobiology, University of Hamburg, University Hospital
Eppendorf, Suderfeldstr. 24, Hamburg D-22529, Germany

FEATURES
source

1.3041
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1..3041
/gene="Stau"
227..1696
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila staufer;
contains four RNA-binding domains; rStau-I6"
/codon_start=1
/product="staufer isoform Stau-I6"
/protein_id="AAf98177.1"
/db_xref="GI:9755163"
/translation="MYKVPDPSRMQSTSYSGMRGAYPPRYFPYPPVLLQVELS
VGGQFNGKMRPVKHDATRALRTLOSEPLPERLEVNGRESEENLNKSEISOV
EIALKRLPNFVARESGPRPHMKNFVTRVSGVEFVGEKSKISKNAARAVLEO
LRLPLPAVERVKPRIKKSQPTCKIQAPDYGQGNPISRLAIOQAKKEPEYM
LITERGLPRREFVMQYVGHHTAEGAGTNKVAKRNAENMLEILGKVPQAQPAK
ALKSEKTPVKKPGDGRKVTFFESPDEGNSNKEDEFMPYLSHOOLPAGILPVP
EVAQAVGSQGHHTKDPARAAPNPAKATVTAMTARELLYGTSPETILKSNISSGH
VPHGPTRPSEQLYLSRAQGFVEYKDFPKNNKNECVSLINCSQPLVSHGIGKDV
ESCHDMAALNLIKLSLELDQOSTEMPRTGNGPVSAACGRG"
polyA_signal
3025..3030
/gene="Stau"

BASE COUNT 814 a 771 c 747 g 707 t 2 others
ORIGIN

Query Match 70.5%; Score 2013.4; DB 74; Length 3041;
Best Local Similarity 87.3%; Pred. No. 0;
Matches 2513; Conservative 0; Mismatches 233; Indels 131; Gaps 23;

QY 91 gcgcctccctctctctagagcccgccgctgcccgcctcgcctcgcctcgcctcgc 150
Db 1 GCGCTCCCCCTTCTTGACCCCGCGCGCGCGCGCGCGCGCTTCTCCGACCACTCCG 60
QY 151 cctctccctctctgctgccccctttctcctgcccgtctcaattgctctcactc 210
Db 61 TCTCTTCCCTCCCGCGT---CTTCTTCTCCTCGCCCTCTTCTTTGCTTTT---CTTC 113
QY 211 ctgcgcgcgcgcgaagaccgcccgcggagcagctctgaggaaagcagccagaagt 270
Db 114 CTGCGCGCGCGCAATACCGCGCGCGCGCGCGCGAGCTCTGGGGAAGCAGCCAGAAAGTA 173
QY 271 tagctctacattgagctcaatgcactgtgtgaaactggaagaaacaaatgtata 330
Db 174 TAGCTTCTACATTGAGCTCAATGACACTGTGCTGAACCTGGAAAGAAACCAATGTATA 233
QY 331 agcccgctgaccctcactctcgatgcactccactacactacatgcatgctgtgagt 390
Db 234 AGCCCGTGACCCCACTCTCGATGACATCCACTACAGCTATGGCATGCGTGAGGTG 293
QY 391 cctatccccccagatactttaccatttcagctccacccttaactctaccagttagc 450
Db 294 CCTATCCCCCAGATTACTTTACCCATTCCAGTTCCACCCGTTACTCTACCAAGTTGAGC 353
QY 451 tctcgcgtggcagcagcaggtttaatggaagaaagaaatgagaccacccgtgaaacag 510
Db 354 TCTCCGTGGGAGGACAGCAGTTTAATGGGAAGGGAAGATGAGACCACTGTGAACAATG 413
QY 511 atgcccctgcccgtgctgaggaactgcaagtgaaacccctgcccagaagaagtgtgag 570
Db 414 ATGCCACTGCCCGAGCCCTGAGGACTCTGAGAGTGAGCCCTTACCAAGAGGCTAGAGG 473
QY 571 taatgtgaagaaagcagaggaagaaacctaataatcgaagaaagcaagtgttg 630
Db 474 TAAATGGAAGAGAAATCAGAAGAAGAAACCTTAATAATCAGAAATTAAGCCAAAGTGTG 533

QY 631 aaattgcgtgaagcggaaatttgctgtgaattttgaggtgccccggagagtgccccac 690
Db 534 AGATTGCGCTGAGCGGAATTGCGTGTGAATTTGAGGTGGCCCCGGAGAGTGGCCCCG 593
QY 691 cacacatgaagaactttgtgaccaggtttcagttgagggaattttgagggaagagaag 750
Db 594 CACACATGAAGAACTTTGTGACCAGGGTTTCAAGTTGGGGAGTTTGTAGGGGAAGGACAAG 653
QY 751 ggaagaaagcaagaatctccaagaagaatgcggccagggctgtcttgagcagcttaga 810
Db 654 GGAAGCAAGCAAGAGATCTCCAAGAAAGATGCCGCAAGGGCCGTGCTGAGCAGCTAAGGA 713
QY 811 ggtgcacccctccctgctgtgtgagcagtgtaagcccagaatccaagaagaagaagtcagc 870
Db 714 GGCTGCCACCTTAACCTGCTGTGAGCGAGTGAAGCCAGAAATCAAGAAAGATGACG 773
QY 871 ccaacctgca-----agacagccccgattatgccaaggtgaatcctattagtagac 924
Db 774 CCACCTGCAAGCTACAGACTGCCCCCGCACTATGGCCCAAGGATGAATCCTATTAGCAGAC 833
QY 925 ttgcacagatccacgcaagcaaaaaaagaaagaaagcagagttacatgtctccttacagaac 984
Db 834 TGGCACAGATCCACGACGCAAAAAAGGAGAAAGAGAGCCGGAGTACATGCTCCTTACAGAAC 893
QY 985 gaggtcttccagctcgcaaggagtttgtgatagcaggtlaaagtttgggcatcacactgag 1044
Db 894 GAGGCTCTCCACGCTGCAGGGAGTTGTGATGCAGGTGAAGGTTGGGCATCACACCCGACG 953
QY 1045 aaggagtggtaccataaagaagtgccaaagcgtlaatgtctgtgagaacatgtctgaga 1104
Db 954 AAGGAGCAGGTACCAATAAAGAAGGTGGCCAAAGCCAAATGCTGTGAGAACATGCTGGAGA 1013
QY 1105 tccctgggttcaaaagttccccagggcgcagcctgccaagccagcactcaatcagaagaga 1164
Db 1014 TCCTGGGGTTCAAGTCCCGCAGCGCAGCCTGCAAGCCAGCACTCAAAATCAGAAAGAGA 1073
QY 1165 agactccagtaagaanaaccagagagcaggaagaaagtaacgtttttgaacctagccctg 1224
Db 1074 AGACTCCAGTAAGAANAACCAGAGAGATGGAAGAAAGTAACATTTTGAACCTAGCCCTG 1133
QY 1225 gggatgaaatgaaactagtaacaagagcagaggttcagagatccttatcttagccatc 1284
Db 1134 GGGATGAATAATGAACCAAGTAATAAGAGGATGAGTTCAAGGATGCCCTATCTTAGCCATC 1193
QY 1285 agcagctgccagctggaattctccccaatggtccggaagttgccaggtgtgcgggtta 1344
Db 1194 AGCAGCTGCCAGCGCGGATTTCTCCCAAGTGTGCCAGAGGTGCCAGCGCTGTGCGGGTTA 1253
QY 1345 gtcaaggaaccacacccaagaatttcaccagggcagctcccaatcctgccaaggaagc 1404
Db 1254 GTCAAGGACACCAATACCAAGAATTTCCGCGAGGGCAGCTCCAAATCCTGCCAAGGCCACG 1313
QY 1405 taactgcatgtagcccgagagtttgttacggygggcaacctgcgccacagccgagacca 1464
Db 1314 TAACTGCCATGATAGCCCGAGAATTGTTGTAACGGGGGACACCTCGCCACAGCCGAGACCA 1373
QY 1465 tttaagaagtaacatctctcagggccagtaaccccatggaactcgcaactagacccttg 1524
Db 1374 TTTTAAAGAGTAACATCTCTTCAAGGACATGTAACCCATGAGCTTGACACAAGACCTCTTG 1433
QY 1525 agcaactgtactactttccagagcccagagatccaggttgaaatacaagaattttcca 1584
Db 1434 AGCAACTGTACTACTTTCAGAGGCCAGGAGATTCCAGGTTGAATTAACAAGACTTCCCA 1493
QY 1585 agaacaacaagaagagtggtatctcatcaactgctcctcaagccgctctgtca 1644
Db 1494 AGAACAACAAGAAGAGTGTATCTCTCAATCAACTGCTTTCACAGCCACCTCTGCTCA 1553
QY 1645 gtcatgcatcggaagagatgtgaggtcctgtcatatgatatgtgtgcaactgaacatttaa 1704
Db 1554 GCCATGGCATCGGCAAGAGATGTGAGAGTCTCTGTATGATATGGCTGCACCTGAACATCTTAA 1613

|||||
Db 1023 TAAAGTTGGGATCAGACTGAGAGAGAGTGGTTACCAATAGAAAGTTGGCCAAAGCGTA 1082
QY 1081 atgctgctgagaacatgctgagatcctggtgttcaaaagttccccagggcagcctgcca 1140
Db 1083 ATGCTGCTGAGAACATGCTGAGATCCTGGGGTTCAAAAGTTCCCGAGGGCAGCCTGCCA 1142
QY 1141 agccagcactcaatcagaagaagaagactccagtaagaagaaccagagagaggaagaag 1200
Db 1143 AGCCAGCACTCAATCAGAGAAGAGACTCCAGTAAGAACAACGAGAGAGGAAGAAAG 1202
QY 1201 taacgttttttgaaacctagccctggtggtatgaaatggaactagtaacaagagcagagt 1260
Db 1203 TAACGTTTTTTGAACCTAGCCCTGGGGATGAAATGGAAGTAAACAAGGACGAGGAGT 1262
QY 1261 tcaggaatgccttatctttagccatcagcagctggcagctggaatctccccatggtgcgg 1320
Db 1263 TCAGGATGCCCTTATCTTAGCCATCAGCAGCTGCCAGCTGGAAATTCCTCCCATGGTGC 1322
QY 1321 aagttgcccagagctgtcggtgttagtcaagagacacacaccaagaatttcaaccagggcag 1380
Db 1323 AAGTTGCCCAGGCTGTGGGGTTAGTCAAGGACACCAACCAAGATTTTCACCAAGGGCAG 1382
QY 1381 ctccaaatcctgccaaggaacggttaactgacctgatatgccccgagagttgtgtacggg 1440
Db 1383 CTCCAAATCCTGCCAAGGCAACGGTAAGTGCATGATAGCCCGAGAGTGTGTACGGGG 1442
QY 1441 gcaacctgcccacagccgaagaccattttaaagagtaacatctctcagccacgtacccc 1500
Db 1443 GCACCTCGCCACAGCCGAGACCATTTTAAAGAGTAACATCTCTTCAGGCCACGTACCCC 1502
QY 1501 atygacctgcactagaccctctgagcaactgtactaccttccagagccaggaattcc 1560
Db 1503 ATGGACCTCGCACTAGACCTCTGAGCAACTGTACTACTCTTCACAGAGCCAGGATTC 1562
QY 1561 aggttgatacaagaattttcccaagaacaacaagaacgagtgttatctctcatcaact 1620
Db 1563 AGGTTGAATACAAAGATTTTCCCAAGAACAACAAGAGAGTGTATCTCTCATCAACT 1622
QY 1621 gctcctcacagccgctctcgctcagtcatgtgcacgcgcaagatgttgagtcctgtcatg 1680
Db 1623 GCTCTCACAGCCGCCCTCGTGCATGTCATGGCATCGGCAAGGATGTGAGTCTCTGCATG 1682
QY 1681 atatggtctgactgaaacattttaaagctgtctgtgagttggaaccaagagcacagaga 1740
Db 1683 ATATGGCTGCACGTGAACATTTTAAAGCTGCTGTCTGAGTTGGACCAACAGACACAGAGA 1742
QY 1741 tgcacaagaacaggaatggaaccagtttcaagctgctcgggaggtgctgaacctttctggcc 1800
Db 1743 TGCCAAGAACAGGAATGAGCCAGTTTCAGCGTCCGGGAGGTGCTGAACCTTTTCTGGCC 1802
QY 1801 acaaaccatlataaaacccaacatatatactgaaataactgagaactgctttgaaattt 1860
Db 1803 ACAAACCATTTATAAAACCCCAACATATATACGTGAATAATCTGAGAACTGCTTTGAAATTT 1862
QY 1861 ggaatatctgataactccagtggtggccaagacatggtgataaaatgtgycaaagacgac 1920
Db 1863 GGAATATCTGATACTCCAGTGGGCCAAGACATGATGATAAAATGTGCCAAAGACGAC 1922
QY 1921 aagaaacttcagtggtgtagccctggtgtgtgctgcggcctagtgatgagtgtgtctgc 1980
Db 1923 AAGAAACTTCAGGTGTAGCCCTGTTGTGTCTGGCGGCTAGTGATGATGCTGTGCTCTGC 1982
QY 1981 catccatccagacagaagaaccagcccccaacgctccagttctgttttgcacgtgacaaa 2040
Db 1983 CATCCATCCAGACAGAAACCAAGCCCAACGCCCTCCAGTTCTGTTTTCATCTGTACAAA 2042
QY 2041 gagagcacagccaattctcatgtgtgcttcttcagataacttgaaaacccgagacgcca 2100
Db 2043 GAGAGCACAGCCAAATCTCATGTGCTCTCTTCAGATACCTTTGAAAAAACCCGAGACGCA 2102
QY 2101 caccagagagccttatagcggccccggagctaaacgaccagagaaaaggccagtgtt 2160
|||||

Db 2103 CACCAAGAGGCCCTTATAGGGCCCCGGAGCTAAACGAGACCAGAGAAAAGGCCAGTGCTT 2162
QY 2161 cctactgcacatgactgactgaagctccgccaacgtagcaccactgttaaccactgttc 2220
Db 2163 CCTACTGCACATGACTGACTAGCTCCGCCACACAGCTAGCACCACTGTAACTACTGCTTTC 2222
QY 2221 tcttc-agtttcatttcttcccttgattgatacaaacactataatttcatctcagttcc 2279
Db 2223 TCTTCAAGTTTCATTTTTCCTTTGATTGATACACACTATATATTTTCATTTCAGTTCC 2282
QY 2280 ttagctgtgttacttacttagcagtttagaaacgtcagtlcatgttaacttgaagatc 2339
Db 2283 TTAGTCGTGTCTACTTACCTAGCAGTTTAGAAACTGTCAATCATGTAACTGCAAGATC 2342
QY 2340 acagcccggttggtgtgcatctctgtccctctgcttgctggaacagttctggaattacca 2399
Db 2343 ACAGCCCGGTTGGGTGGCATTCGTGCTCTGGCTTGAGACAGTCTTGGAATTACCA 2402
QY 2400 ccagaatccttgaactccctgccccctgtataaatggaacagcttaggacttttaaactt 2459
Db 2403 CCAGAACTCTTGACTCCTCGCCCTTGTTATAAATTGGACAGCTTAGGACTTTTAAACTTT 2462
QY 2460 agatcaaaaagatatgtctctttaaacttattttaaaggagcagactttaaagtacc 2519
Db 2463 AGATCAAAAAGATATGTCTCTTTTAACTTATTTTTAAAGAGACAGACTTTAAAGACC 2522
QY 2520 ctgaccttaccatataacagaatttgtcaaaaggagtglttcttgaggagtagctt 2579
Db 2523 CTGACCTTTAACCATATATACAGAAATTTGTCAAAAAGAGAGTGTCTTGAGAGGTTAGCTT 2582
QY 2580 tttttaccacactacagacattacctgttagggccagaagaactacagctgtgtccct 2639
Db 2583 TTTTTTAACCACTACAGGACATTACCTGTAGGCCCAAGAGACTACAGGCTGTCTCCCT 2642
QY 2640 agagggcccaataacagtcgaattccaacctctaagtcggggaaaggtgaagggttctctg 2699
Db 2643 AGAGGGCCCAATACAGTCAATTCCAACCTCTAAGTCGGGGAAAGGTGACAGGTTTCTCTGG 2702
QY 2700 tgcgtgtgtgcacagggggcaggcaggtcagctgctggtgggaagagcatgtgtcccta 2759
Db 2703 TGCTGTGTGCACAGGGGGCAGGTCAGCTGGCCTGGGGAAGAGCATGTGCTCCTA 2762
QY 2760 gtgcagccctgtcttcacactctgtgtttagctggaaccttccacatgtaataagta 2819
Db 2763 GTGCAGCCCTGCTTCACCTTGTGTTAGCTGGAACCTTCCCACTCATGGAATATAAGTA 2822
QY 2820 aactcaacttcttgttcaccaataaatgtaataactaa 2857
Db 2823 AACTCACTTCTTGTGCACCAATAAATGTAATACTAA 2860

RESULT 2
AF290989
LOCUS 3041 bp mRNA ROD 09-AUG-2000
DEFINITION Rattus norvegicus stauven isoform Stau-I6 (Stau) mRNA, complete
ACCESSION AF290989
VERSION AF290989.1 GI:9755162
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3041)
AUTHORS Monshausen,M., Putz,U., Rehbein,M., Schweizer,M.,
DesGroselliers,L., Kuhl,D., Richter,D. and Kindler,S.
TITLE Two rat brain stauven isoforms differentially bind RNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3041)
AUTHORS Monshausen,M. and Kindler,S.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) Institute for Cell Biochemistry and

22	122.8	4.3	2008	34	AK023314	AK023314 Homo sapi
23	120.4	4.2	160979	52	AC055879	AC055879 Homo sapi
C						
24	113.2	4.0	332811	72	AL390198	AL390198 Homo sapi
25	95.8	3.4	5237	53	DROSTAFEN	M69111 D.melanogas
26	95.8	3.4	61267	28	AC004336	AC004336 Drosophil
C						
27	95.8	3.4	120625	38	AC017563	AC017563 Drosophil
C						
28	95.8	3.4	267271	28	AE003800	AE003800 Drosophil
29	95.6	3.3	8002	29	AF225925	AF225925 Musca dom
30	91.2	3.2	8409	29	AF225924	AF225924 Drosophil
C						
31	87	3.0	171591	50	AC027018	AC027018 Homo sapi
32	70.6	2.5	2847	69	HSRNATTFI	X83973 H.sapiens m
33	65	2.3	1166	69	HST000002	AL078636 Homo sapi
34	64.2	2.2	7218	81	I66494	I66494 Sequence 14
C						
35	64	2.2	7218	81	I66494	I66494 Sequence 14
C						
36	57.8	2.0	42099	29	CELF55A4	U67949 Caenorhadi
C						
37	54.6	1.9	212493	62	AC080167	AC080167 Mus muscu
C						
38	53.8	1.9	93821	46	AC021596	AC021596 Homo sapi
C						
39	52.8	1.8	87120	36	AC012225	AC012225 Homo sapi
C						
40	52.6	1.8	78690	38	AC016541	AC016541 Homo sapi
41	52.6	1.8	164520	39	AC020738	AC020738 Homo sapi
42	52.2	1.8	61363	54	AC068364	AC068364 Homo sapi
C						
43	52.2	1.8	280047	54	AC068494	AC068494 Mus muscu
44	52	1.8	58514	62	AC079898	AC079898 Homo sapi
45	52	1.8	64272	47	AC023019	AC023019 Homo sapi

ALIGNMENTS

RESULT	1
AF061942	
LOCUS	
DEFINITION	AF061942 2860 bp mRNA ROD 04-MAR-1999
ACCESSION	Mus musculus RNA-binding protein stauflen mRNA, complete cds.
VERSION	AF061942
KEYWORDS	AF061942.1 GI:4335944
SOURCE	.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 2860)
TITLE	Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroseillers,L. Mammalian stauflen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum
JOURNAL	Mol. Cell Biol. 19 (3), 2220-2230-(1999)
MEDLINE	99147057
REFERENCE	2 (bases 1 to 2860)
AUTHORS	Duchaine,T., Luo,M. and DesGroseillers,L.
TITLE	Direct Submission
JOURNAL	Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES	Location/Qualifiers
source	1..2860

```

CDS
325. .1789
/codon_start=1
/product="RNA-binding protein staufen"
/protein_id="AAD17529.1"
/db_xref="GI:4335945"
/translation="MYKPVDPHRSRMOSTYSYGMRGAYPRFYFPFVPLLQVELSL
VGGOQFNGKGMKRPVVKHDAFARALRLQSEPLPERLEVNGREAEEENLNKSEISOVF
EIALKRNLPVNFEEVARESGPPIHMKNLPTRVASGEFVGEEGSKISKNAARAVLEQ
LRLPLPLPAVERVKPRIRKKKSQPTCTAPDYGQGNPISRLAQIOAKKEKEPEYMLL
TERGLPRRRFVMOVKVGHTTAEGVGNKKVAKRNAEENMLELIGFVYQOAPAKPALI
KSEKTPVKKKPGDGRKVTFFEPSPGDENGTSNKDEEFRMPLYLSHQQLPAGILMPVEV
AOAVGSQGHHTKDTFRAAPNPAKATAMIARELLYGGTSPTAETLISNLSGSHVP
HGPRTRPSEOLYYLSRAQGFQVEYKDFPKNNKNECVSLINCSSOPLVYSHCIQKDVES
CHDMAALNLIKLSLSDQOSTEMPTGNGPVSACGRC"

```

Query Match	99.6%;	Score 2846;	DB 74;	Length 2860;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2857;	Conservative	0;	Mismatches	0;
			Indels	1;
			Gaps	
QY	1	cggcgcgccctcccccgtcactctcgtccaggtcgggcccagccgctctcaag	60	
Db	3	CGGGCGGCCCCCTCCCGCTCACCTTCCGACGAGGTGCGGGCCCGACCGCTTTCAG	62	
QY	61	cgttgcgtgctgtgcgtctgtgtgtgcgtcccccctctctctagccccgcctg	120	
Db	63	CGTTTGGCGTGGCTGTGCTGCGGTGTGTGTGGCTCCCGCTTCTTGTAGCCCCGGCCTG	122	
QY	121	gcgcgcgcgccttcgcctccgcacatccgcctctccctcctctgtgcgccctttt	180	
Db	123	GGGGGCCCCCTTCGCTCCGCCACATCCGCCCTTCCCTCCCTGTGGTGGCCCCCTTTT	182	
QY	181	cctgcgcgtctcacttgcctctctcactctctcgcgcgcgcgcgcgcgcgcgcgcg	240	
Db	183	CCTGCGCGTCTTCACTTGCTTCTTCACTTCTCGCGCGCGCCCAAGACCGCGGCCCGG	242	
QY	241	gacgagctctggggaaagcagccagaaagataagctctacatctagctcaatgcactgt	300	
Db	243	GACGAGCTCTGGGGAAGCAGCCAGAAAGATAGCTTCTACCATATGAGCTCAATGCACGT	302	
QY	301	gtgtgaaactggaagaaacaacatgtataagcccggtggaacctcaactcgtgattgact	360	
Db	303	GTGTGAAACTGGAAGAAAGAAACCAATGTATAAGCCCGTGGAACCTCACTCTCGATGCAGT	362	
QY	361	ccacctacagctatgcatgctgtgaggtgctatccccccagatactttaccatttc	420	
Db	363	CCACCTACAGCTATGGCATGCGTGGAGTGGCTATCCCCCAGATFACCTTTACCCATTTC	422	
QY	421	cagtcaccaacttactctaccaaagtttagctctccgttggcgagcagcagtttaatygga	480	
Db	423	CAGTCCCACTTACTCTACCAAGTTGAGCTCTCCGTGGGGGAGACAGCTTAATGGGA	482	
QY	481	aaggaagaatgagaaaccaaccggtgaacaacgatagccctgcgcgtgagactctgc	540	
Db	483	AAGGAAAGATGAGACCAACCCTGTAACACGATGCCCTTGCCTGTGAGGACTCTGTC	542	
QY	541	agagtgaaacccctgcagaaaggttggaggtaaalvgaaagaaagcagaggaagaaaac	600	
Db	543	AGAGTGAACCCCTCGCAGAAAGGTTGAGGTTAAATGGAAGAGAGCAGAGAAAGAAACC	602	
QY	601	tcaataaatcggaataaagccaagtgtttgaaatgtgcgtgaagcggaaattgacctgtga	660	
Db	603	TCATAATAATCGGAATAAAGCCAAAGTGTGTAATAATTGGCTGAAGCGGAATTTGCCGTGA	662	
QY	661	attttgagtgccccgggagagtggtgcccacacacatgaagaactttgtgaccaggttc	720	
Db	663	ATTTTGAGGTGGCCCCGGGAGAGTGGCCCCACACACATGAAGAACTTGTGACCAAGGGTTT	722	
QY	721	cagttggggaattgttaggggaaagagaaagggaaaaagcaagaagaatctccaagaagaatg	780	
Db	723	CAGTTGGGAATTTGTAGGGGAAGGAGAGGAAGGAAGCAAGAAGATCTCCAAGAAGATG	782	
QY	781	cggccagaggtgttcttgyagcagcttagagggctgccaacccctccctgttlygaacgag	840	
Db	783	CGGCCAGGGCTGTCTTGAGAGCAGCTTAGGAGGTGCCACCCCTCCCTGCTGTGGAGCGAG	842	
QY	841	tgaagcccaagaatcaagaagaaagaatcagcccaactgcaagacagccccgattatggcc	900	
Db	843	TGAAGCCCAGAAATCAGAAGAAAGTCAAGCCCACTGCAAGACAGCCCCGGATTATGGCC	902	
QY	901	aaggatgaatctctatagtagacttgcacagatccagcagggcaaaaaagagaagagac	960	
Db	903	AAGGATGAATCCTATTAGTAGACTTGCACAGATCCAGACGCAAAAAGAGAAGAGAGC	962	
QY	961	cagaatcatgtctccttaacagaacgaggtcttccaagctgcgaaggagtttgtatgcagg	1020	
Db	963	CAGAGTACATGCTCTTACAGAACGAGGCTTTCACAGCTGCGCAGGGAGTTGTGATGCAGG	1022	
QY	1021	taaaagttgggcatacactgcagaagaaggtggttaaccaataagaaggttggccaagcgt	1080	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 09:22:49 ; Search time 24649.1 Seconds
(without alignments)
593.183 Million cell updates/sec

Title: US-09-316-048-9
Perfect score: 2857
Sequence: 1 cgggcgcgcgcgcctcccccgcg.....ccaataatgtaataactaa 2857

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_bal:*
 - 2: gb_ba2:*
 - 3: gb_om:*
 - 4: gb_ov:*
 - 5: gb_ph:*
 - 6: gb_pil:*
 - 7: gb_pl2:*
 - 8: gb_pr1:*
 - 9: gb_pr2:*
 - 10: gb_pr3:*
 - 11: em_fun:*
 - 12: em_hum1:*
 - 13: em_hum2:*
 - 14: em_in:*
 - 15: em_om:*
 - 16: em_or:*
 - 17: em_ov:*
 - 18: em_pat:*
 - 19: em_ph:*
 - 20: em_pl:*
 - 21: em_ro:*
 - 22: em_sts:*
 - 23: em_sy:*
 - 24: em_un:*
 - 25: em_vi:*
 - 26: gb_htg1:*
 - 27: gb_htg2:*
 - 28: gb_in1:*
 - 29: gb_in2:*
 - 30: em_bal:*
 - 31: em_ba2:*
 - 32: em_hum3:*
 - 33: em_hum4:*
 - 34: gb_pr4:*
 - 35: gb_htg3:*
 - 36: gb_htg4:*
 - 37: gb_htg5:*
 - 38: gb_htg6:*
 - 39: gb_htg7:*
 - 40: em_htg1:*
 - 41: em_htg2:*
 - 42: em_htg3:*
 - 43: em_hum5:*

- 44: gb_pl3:*
- 45: gb_pr5:*
- 46: gb_htg8:*
- 47: gb_htg9:*
- 48: gb_htg10:*
- 49: gb_htg11:*
- 50: gb_htg12:*
- 51: gb_htg13:*
- 52: gb_htg14:*
- 53: gb_in3:*
- 54: gb_htg15:*
- 55: gb_htg16:*
- 56: gb_htg17:*
- 57: em_htg4:*
- 58: em_htg5:*
- 59: em_htg6:*
- 60: em_htg7:*
- 61: em_hum6:*
- 62: gb_htg18:*
- 63: gb_htg19:*
- 64: gb_ba3:*
- 65: em_htg8:*
- 66: em_htg9:*
- 67: em_htg10:*
- 68: gb_pr6:*
- 69: gb_pr7:*
- 70: gb_htg20:*
- 71: gb_htg21:*
- 72: gb_htg22:*
- 73: gb_htg23:*
- 74: gb_ro:*
- 75: gb_scs1:*
- 76: gb_sts2:*
- 77: gb_sy:*
- 78: gb_un:*
- 79: gb_vil:*
- 80: gb_viz:*
- 81: gb_pat1:*
- 82: gb_pat2:*
- 83: em_htg0:*
- 84: gb_htg24:*
- 85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query		DB	ID	Description
No.	Score	Match	Length			
1	2846	99.6	2860	74	AF061942	AF061942 Mus muscu
2	2013.4	70.5	3041	74	AF290989	AF290989 Rattus no
3	1985.4	69.5	3059	74	AF227200	AF227200 Rattus no
4	1354.6	47.4	3142	10	AF061940	AF061940 Homo sapi
5	1269.8	44.4	3217	10	AF061938	AF061938 Homo sapi
6	1256	44.0	3506	10	AF061939	AF061939 Homo sapi
7	1255	43.9	3349	10	AF061941	AF061941 Homo sapi
8	1253	43.9	3194	34	AK025519	AK025519 Homo sapi
9	1209.2	42.3	3066	68	HS132258	AJ132258 Homo sapi
10	1045.2	36.6	225016	52	AC063967	AC063967 Mus muscu
11	325.4	11.4	360	74	RNO010200	AJ010200 Rattus no
12	283.6	9.9	113836	34	AL133174	AL133174 Human DNA
13	279.4	9.8	2968	34	AK002152	AK002152 Homo sapi
14	279.4	9.8	4058	68	HS132258	Y19062 Homo sapien
15	210.2	7.4	163253	68	HSB140E4	AL121877 Human DNA
16	188.4	6.6	7241	74	RNCALBD9	X16635 Rattus norv
17	159	5.6	160979	52	AC055879	AC055879 Homo sapi
18	158.2	5.5	194407	54	AC068845	AC068845 Homo sapi
19	156.6	5.5	157803	35	AC011492	AC011492 Homo sapi
20	154.6	5.4	3173	34	AK001576	AK001576 Homo sapi
21	127.4	4.5	1300	74	RATCBBD9K	M83775 Rat calbind

Db 205 TGTAAATACATCTCCATTAGGGCTGAAAGAATGACCTACGTTTCTGTATACAGCTGTG 146

QY 3151 ttgcttttgatgtgtgttactgttacacagaagtgtgtgcactgaggtcttgcgtgtgt 3210

Db 145 TTGCTTTTGATGTGTGTACTGTACACAGAAGTGTGTGCACGTAGGCTCTGCGTGTGCT 86

QY 3211 ccgtatggaagaacctgtgtagccctgcgaglttaagtactgtcttccattcattgttaacgct 3270

Db 85 CCGTATGGAAAGCCGTGTAGCCCTGCGAGTTAAGTACTGCTTCATTCATTGTTTACGCT 26

QY 3271 ggaattttctcccatgtaatgta 3295

Db 25 GGAATTTTCTCCCATGAAATGTA 1

RESULT 15

RNO010200 360 bp mRNA ROD 22-JAN-1999

LOCUS Rattus norvegicus mRNA for Staufen protein, partial.

DEFINITION AJ010200

ACCESSION AJ010200.1 GI:4138427

VERSION staufen gene.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 360)

AUTHORS Kiebler,M.A.

TITLE Direct Submission

JOURNAL Submitted (16-AUG-1998) Kiebler M.A., Cell Biology, EMBL, Meyerhofstr. 1, D-69117 Heidelberg, GERMANY

REFERENCE 2 (bases 1 to 360)

AUTHORS Kiebler,M.A., Hemraj,I., Verkade,P., Kohrmann,M., Fortes,P., Marion,R.M., Ortin,J. and Dotli,C.G.

TITLE The mammalian staufen protein localizes to the somatodendritic domain of cultured hippocampal neurons: implications for its involvement in mRNA transport

JOURNAL J. Neurosci. 19 (1), 288-297 (1999)

MEDLINE 99088098

FEATURES

source Location/Qualifiers

1..360

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone_lib="Rat brain"

/dev_stage="embryonic"

1..119

/gene="staufen"

<1..>119

/gene="staufen"

/codon_start=2

/evidence=experimental

/product="Staufen protein"

/protein_id="CAA09037.1"

/db_xref="GI:4138428"

/translation="YFYPFPVPLLYQVELSVGGQGFNGKMKRPPVKHDATA"

BASE COUNT 109 a 70 c 105 g 76 t

ORIGIN

Query Match 8.6%; Score 287; DB 74; Length 360;

Best Local Similarity 87.5%; Pred. No. 2,1e-57;

Matches 314; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 596 tactttaccatttcagttccaacttacttatacaagtgaacttctgtggagaa 655

Db 2 TACTTTTACCATTTCAGTTCACCGTTACTTACCAAGTTGAGCTCTCCGTGGAGGA 61

QY 656 cagcaatttaatgycagaagaagaacaagacagcgtcgcaacaacgatgtgtgccaac 715

Db 62 CAGCAGTTTAATGGGAAGGAAGATGAGACCACCTGTGAACAATGATGCCACGTGCCCGA 121

QY 716 gcgttgagatcctgcagatgagccctgtccagagaggtgaggtgaatggaagaa 775

11 111111 1111111 11111111 111111111 11111 111111111111

Db 122 GCCCTGAGGACTCTGCAGAGTGAGCCCCCTACCAGAGAGGCTAGAGTTAAATGGAAGAA 181

QY 776 tccgaagaagaanaatctcaataaatctgaanaatgaagtgatgtgacttaaa 835

Db 182 TCAGAGAAGAAACCTTAATTAATCAGAATAAAGCCAAAGTGTGAGATTGCCCTGAAG 241

QY 836 cgaacttgccctgtgaatttcgaagltgcccgggagaagtgcccaaccacatgaagaac 895

Db 242 CGGAATTTGCCCTGTGAATTTGAGGTGGCCCGGAGAGTGGCCCGCACACATGAAGAAC 301

QY 896 ttgtgaccaaggttctcglttgaggagltgtgtggggaagtgaaaggaagaagaac 954

Db 302 TTTGTGACCAGGCTTTCAGTTGCGGAGTTTGTAGGGGAAGAGAAGGAAAGCAAGAA 360

Search completed: April 5, 2001, 09:22:49

Job time: 80614 sec

Best Local Similarity 73.2%; Pred. No. 9.7e-80;
Matches 586; Conservative 0; Mismatches 203; Indels 11; Gaps 7;

QY	2308	gtctttttttccctgtgtgaaagaacgacgaccccccttcaagctgtctaac	2367
DB	60919	GTATTTTATTTGGCTGTGTGAAAAAAGAAATGGAGACAGAAAGCATTACCCGTCTCAC	60860
QY	2368	tcagacacattggyacaacacccctgagccatgcccagagagagcccttgaccgcccc	2427
DB	60859	ACTGGCTGACTCAAC-ACCTTGCTATTAACATGCTGGAGAAAAACCTTAGACTGGCTTC	60801
QY	2428	agagctaaagcaccagagaaatcaaatgcttctactcagcgtgaccacatttcta	2487
DB	60800	AAAGCTGAAAAACACCGGGGAAATTAACCTCTTCTACTCGTTACAACATTACCTTCTTA	60741
QY	2488	gtgtgccagggcccccacaccctcctgagtaaccacacacacacacacgtcttcttc	2547
DB	60740	GCATGCTACGG-CCAGTCACTTACTGCACTACCCACACCATACCACTGCTTCTCTCT	60682
QY	2548	aacagtgtatctgtatctttagtttcttcttcttcttcttcttcttcttcttcttct	2607
DB	60681	AATAGTGATACATATTTTATTTAGTTTATTA-TTCTTTTGATTTGAATGATGCCATATGAA	60623
QY	2608	atttctattgagaattctcctcaattgatactagtttaaatagcacagtttggaaactgtc	2667
DB	60622	ATTTCATTTGAAAGTTTCTCAGTTGATCTAGTTACATAGTACAGTTTCAGAAACCTGTC	60563
QY	2668	tgagactgacttatacaataatcaaccgcaaaagatcatatcatatgtatgtgtag	2727
DB	60562	CCAGACTGACTTTATCAGGAATCTAACCTGAGCATAGATCATATCATATGATACGTGTGT	60503
QY	2728	acattttatttctactgactaacaccagacagtttcaagtgtatgcaaatgtgtgcccc	2787
DB	60502	ACATTCTTTTCTACTGACTAACAAGAACCATTTTCAGTGATGCAAAATCATGTGACCTCT	60443
QY	2788	ggttcagctgaaacagtcctgactt-tcaaaaacaccttgaaatagtcctccacagttgta	2846
DB	60442	GGTTTAGCTGAAAAAGCCCTGAGACTTTCAGAGAAACCTTGATCAAAATTTCCACTATGTGA	60383
QY	2847	taatttgacaatttaggaattttaaactttagatgatactattggttccattttatttc	2906
DB	60382	CAATTTGGCCATTTTAGGAATTTCAAACTTCA---ATCATTTGGTCTCTTTTATATTTT	60327
QY	2907	attttatttttctttaaataatgca--aacagagacttaaatgaacttgcctctgttttaag	2964
DB	60326	ATTTTATTTTGGTTGTGTAATGGGACCTTAACAATAAATTAACCTTGATCTTTGTTTAAAG	60267
QY	2965	attataaaaaaacattgtgtatcatatcatatgctcttgaggacttagcttctactaca	3024
DB	60266	ATTATTAATAAACTTACGTGTATGTACATPAGCTCTTCAAGACCCACAGCTTTCACATAG	60207
QY	3025	ctacagatatgatactccatgtagtccatataaaccctgcagagtgatttcc-agagtg	3083
DB	60206	CTGTAGGATATTTGCTCTCTGTAGTACATATGACATTGCAGGTTAATTTTCTTGAGTGC	60147
QY	3084	tcgatactgttaattacatc 3103	
DB	60146	TTTATACTCTTATTTTCTC 60127	

RESULT 14
LOCUS G30939/c 385 bp DNA STS 29-SEP-1998
DEFINITION DRES50 Human DRES Homo sapiens STS genomic clone 22368 3', sequence tagged site.
ACCESSION G30939
VERSION G30939.1 GI:1724032
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 385)

AUTHORS Banfi,S., Borsani,G., Rossi,E., Bernard,L., Guffanti,A.,
Rubboli,F., Marchitelli,A., Giglio,S., Coluccia,E., Zollo,M.,
Zuffardi,O. and Ballabio,A.
TITLE Identification and mapping human cDNAs homologous to Drosophila
JOURNAL mutant genes through EST database searching
MEDLINE Nat. Genet. 13 (2), 167-174 (1996)
COMMENT 96225443

Contact: Ballabio Andrea
Telethon Institute of Genetics and Medicine, TIGEM
Telethon Institute of Genetics and Medicine, TIGEM
via Olgettina 58, 20132 Milano, Italy
Tel: +39-2-215601
Fax: +39-2-21560220
Email: dres@tigem.it
Primer A: GTAACAATGAATGGAACAG
Primer B: GGGCTGAAAAAGATGACCTAC
STS size: 157
PCR Profile:
Presoak: 94 degrees C for 2 minutes
Denaturation: 94 degrees C for 45 seconds
Annealing: 57 degrees C for 45 seconds
Polymerization: 72 degrees C for 45 seconds
PCR Cycles: 35
Thermal Cycler: MJR PTC-200

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 220 uM
Taq Polymerase: 0.025 units/uL
Total Vol: 15 uL
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

The 5' end of the cDNA clone 22368 is homologous to the staufer
Drosophila gene product. EST from which STS is derived: T89105.
Forward primer is TO-1233; Reverse primer is TO-1234.
FEATURES
Location/Qualifiers
1..385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="20, 12.10 cR from GCT10F11 (lod >3.0)"
/clone="22368"
/clone_lib="Human DRES"
STS
29..185
primer_bind 29..49
primer_bind complement(165..185)
BASE COUNT 135 a 79 c 71 g 100 t
ORIGIN

Query Match 11.4%; Score 380.2; DB 75; Length 385;
Best Local Similarity 99.2%; Pred. No. 2.5e-79;
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	2911	ttattttgttaatgcaaacagagacttaaatgaacttgcctctgttttaagattatt	2970
DB	385	TTTATTTTGTAAATGCAACAGAGACTTAATGAACCTTGATCTCTGTTTAAAGATTATT	326
QY	2971	aaaaaacattgtgtatcatatcatatgctcttgaggacttagcttccactacacag	3030
DB	325	AAAAACATTGTGTAATCTATACATATAGGCTCTTGAGGACTTAGCTTTCACTACACTACAG	266
QY	3031	gatatgctcctcatgtagtccatataaacctgcagagtgatttccagagtgctcgatac	3090
DB	265	GATATGATCTCCATGTAGTCCATATTAACCTGCAGAGTGATTTCCAGAGTGCCTGATAC	206
QY	3091	tgtaattacatctccattagggctgaaagaatgacctacgttctgtatacagctgtg	3150

```
QY 2182 tgatggcactcgtgtacctggcggtccctacgcaatagcagctgctgtgggaagaag 2241
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3006 tgacggccactcgtgacctggcgtgtccctacgcaataccagctgcttggggaagaag 2947
QY 2242 ggcggccagccagctgtgtctcccgagacaccagcagatccacaccctgggcacctcg 2301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2946 ggcggccagccagctggtgtctcctggggcacccagatccacgccctgggcaccttcc 2887
QY 2302 tgttggctcttttttccctgtg--tgaagaagaagaacggcagcagcccttctcaagc 2359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2886 tgtttggtcttttttcccggtgtgaagaagaactgaagaagcattgaccttctcaagc 2827
QY 2360 tggctcactcagacacatggygacaacacctgycagccatgcccagagaagggccttga 2419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2826 tggcctta-----gttgggacaaacccctggacagccacacacagagagagcccttcaa 2776
QY 2420 ccggcccccagagctaaagcacaccagagaanaatcaatgcttctactcagcgtgacccaa 2479
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2775 ccagcccccagagctaaagcacaccagagaagaatgcttctctatgacgacaccttaa 2716
QY 2480 ctcttctagtgtgcacggcccccacacctctgtagtaccacacacacacacctgctt 2539
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2715 c--ttctagtggtgcacggcccccacacctctgtagtaccacacacacacacctgctt 2658
QY 2540 tctcttccacagtgatcgtatcttctagtcttcatatttcttcttgatgatatgacac 2599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2657 tctcttccacagtgatcgtatcttctagtcttcatatttcttcttgatgatatgacac 2598
QY 2600 tatataaaatttcat 2615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2597 tatataaaatttctt 2582
```

```
RESULT 13
AL158850/c
LOCUS      AL158850      192597 bp      DNA      HTG      06-SEP-2000
DEFINITION Homo sapiens chromosome 6 clone RP3-522B19, *** SEQUENCING IN
PROGRESS ***, 11 unordered pieces.
ACCESSION  AL158850
VERSION    AL158850.4  GI:9800143
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 192597)
AUTHORS    Burton,J.
TITLE      Direct Submission
JOURNAL    Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Aug 13, 2000 this sequence version replaced g1:9212635.
COMMENT    ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            ----- Project Information
            Center project name: dj522B19
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 12% of reads
            Sequencing vector: plasmid; L08752; 87% of reads
            Chemistry: Dye-terminator Big Dye; 77% of reads
            Chemistry: Dye-primer-amersham; 22% of reads
            Chemistry: Dye-primer Big Dye; 0% of reads
            Consensus quality: 185625 bases at least Q40
            Consensus quality: 188533 bases at least Q30
            Consensus quality: 190005 bases at least Q20
            Insert size: 191597; sum-of-contigs
            Insert size: 144116; 9.6% error; agarose-ff
            Quality coverage: 5.71x in Q20 bases; sum-of-contigs Quality
```

```
coverage: 8.21x in Q20 bases; agarose-ff
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 13837: contig of 13837 bp in length
* 13838 13937: gap of 100 bp
* 13938 17476: contig of 3539 bp in length
* 17477 17576: gap of 100 bp
* 17577 20005: contig of 2429 bp in length
* 20006 20105: gap of 100 bp
* 20106 32070: contig of 11965 bp in length
* 32071 32170: gap of 100 bp
* 32171 39772: contig of 7602 bp in length
* 39773 39872: gap of 100 bp
* 39873 43471: contig of 3599 bp in length
* 43472 43571: gap of 100 bp
* 43572 51795: contig of 8224 bp in length
* 51796 51895: gap of 100 bp
* 51896 178838: contig of 126943 bp in length
* 178839 178938: gap of 100 bp
* 178939 181919: contig of 2981 bp in length
* 181920 182019: gap of 100 bp
* 182020 184231: contig of 2212 bp in length
* 184232 184331: gap of 100 bp
* 184332 192597: contig of 8266 bp in length.
Location/Qualifiers
1.192597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-522B19"
/clone_1lb="RPC1-3"
1.13837
/note="assembly_fragment:02383
fragment_chain:1"
13938.17476
/note="assembly_fragment:01080
fragment_chain:1"
17577.20005
/note="assembly_fragment:02434
fragment_chain:1"
20106.32070
/note="assembly_fragment:03060
fragment_chain:2"
32171.39772
/note="assembly_fragment:00945
fragment_chain:2"
39873.43471
/note="assembly_fragment:00013
fragment_chain:2"
43572.51795
/note="assembly_fragment:02225
fragment_chain:3"
51896.178838
/note="assembly_fragment:01767
fragment_chain:3"
178939.181919
/note="assembly_fragment:01637"
182020.184231
/note="assembly_fragment:02408"
184332.192597
/note="assembly_fragment:02960"
39515 c 38313 g 55431 t 1014 others
```

```
BASE COUNT 58324 a 39515 c 38313 g 55431 t 1014 others
ORIGIN
Query Match 11.4%; Score 383.2; DB 70; Length 192597;
```



```
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1755: contig of 1755 bp in length
* 1756 1855: gap of 100 bp
* 1856 3429: contig of 1574 bp in length
* 3430 3529: gap of 100 bp
* 3530 5346: contig of 1817 bp in length
* 5347 5446: gap of 100 bp
* 5447 7338: contig of 1892 bp in length
* 7339 7438: gap of 100 bp
* 7439 10297: contig of 2859 bp in length
* 10298 10397: gap of 100 bp
* 10398 13590: contig of 3193 bp in length
* 13591 13690: gap of 100 bp
* 13691 16293: contig of 2603 bp in length
* 16294 16393: gap of 100 bp
* 16394 19065: contig of 2672 bp in length
* 19066 19165: gap of 100 bp
* 19166 22713: contig of 3548 bp in length
* 22714 22813: gap of 100 bp
* 22814 27165: contig of 4352 bp in length
* 27166 27265: gap of 100 bp
* 27266 32377: contig of 5112 bp in length
* 32378 32477: gap of 100 bp
* 32478 36572: contig of 4095 bp in length
* 36573 36672: gap of 100 bp
* 36673 41657: contig of 4985 bp in length
* 41658 41757: gap of 100 bp
* 41758 47024: contig of 5267 bp in length
* 47025 47124: gap of 100 bp
* 47125 52938: contig of 5814 bp in length
* 52939 53038: gap of 100 bp
* 53039 59129: contig of 6091 bp in length
* 59130 59229: gap of 100 bp
* 59230 65896: contig of 6667 bp in length
* 65897 65996: gap of 100 bp
* 65997 89427: contig of 23431 bp in length
* 89428 89527: gap of 100 bp
* 89528 95785: contig of 6258 bp in length
* 95786 95885: gap of 100 bp
* 95886 102762: contig of 6877 bp in length
* 102763 102862: gap of 100 bp
* 102863 110899: contig of 8037 bp in length
* 110900 110999: gap of 100 bp
* 111000 121002: contig of 10003 bp in length
* 121003 121102: gap of 100 bp
* 121103 133048: contig of 11946 bp in length
* 133049 133148: gap of 100 bp
* 133149 145755: contig of 12607 bp in length
* 145756 145855: gap of 100 bp
* 145856 165006: contig of 19151 bp in length
* 165007 165106: gap of 100 bp
* 165107 188243: contig of 23137 bp in length
* 188244 188344: gap of 100 bp
* 188344 194407: contig of 6064 bp in length.
*
FEATURES
Source
1. 194407
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19"
/clone="RP11-73711"
/clone_lib="RPC1-11 Human Male BAC"
1. 1755
/misc_feature
/feature="assembly_fragment"
1856. 3429
/misc_feature
/feature="assembly_fragment"
3530. 5346
/misc_feature
/feature="assembly_fragment"
5447. 7338
/misc_feature
/feature="assembly_fragment"
7439. 10297
/misc_feature
/feature="assembly_fragment"
10398. 13590
/misc_feature
/feature="assembly_fragment"
13691. 16293
/misc_feature
/feature="assembly_fragment"
16394. 19065
/misc_feature
/feature="assembly_fragment"
19166. 22713
/misc_feature
/feature="assembly_fragment"
22814. 27165
/misc_feature
/feature="assembly_fragment"
27266. 32377
/misc_feature
/feature="assembly_fragment"
32478. 36572
/misc_feature
/feature="assembly_fragment"
36673. 41657
/misc_feature
/feature="assembly_fragment"
41758. 47024
/misc_feature
/feature="assembly_fragment"
47125. 52938
/misc_feature
/feature="assembly_fragment"
53039. 59129
/misc_feature
/feature="assembly_fragment"
59230. 65896
/misc_feature
/feature="assembly_fragment"
65997. 89427
/misc_feature
/feature="assembly_fragment"
89528. 95785
/misc_feature
/feature="assembly_fragment"
95886. 102762
/misc_feature
/feature="assembly_fragment"
102863. 110899
/misc_feature
/feature="assembly_fragment"
111000. 121002
/misc_feature
/feature="assembly_fragment"
121103. 133048
/misc_feature
/feature="assembly_fragment"
133149. 145755
/misc_feature
/feature="assembly_fragment"
145856. 165006
/misc_feature
/feature="assembly_fragment"
165107. 188243
/misc_feature
/feature="assembly_fragment"
188344. 194407
/misc_feature
/feature="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 50993 a 43965 c 46383 g 50458 t 2608 others
ORIGIN
Query Match 16.5%; Score 553.2; DB 54; Length 194407;
Best Local Similarity 90.4%; Pred. No. 9.6e-120;
Matches 665; Conservative 0; Mismatches 48; Indels 23; Gaps 6;
QY 1883 gctgcgtgaacatcttaagtgtgtctgtgagtgtagtgcacacaaagtacagatgcca 1942
Db 3297 GCTGCACCTGAACACCTTAAAGTTGCTGTGAGTTGACCAACAAAGTACAGATGCCA 3238
QY 1943 agaacagaaacgacccaatgtctgtgtgtgagtgtagtgcacacaaagtacagatgcca 2002
Db 3237 AGAACAGGAATGATCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGGCCATGCAA 3178
QY 2003 ccattataaaatcccaacatatatactgaaataactgaaactgcttgaatattggaat 2062
Db 3177 CCATTATAAAATCCCAACATATATACACTGAAATAATTAATAAACTGCTTTGAAAATTGGAAT 3118
QY 2063 ttctgatacctccagtggtggtgagagagacagtggtgtaagagatgtggtgcagcagcagg 2122
Db 3117 TTCTGATACCTCTAGTGGCCGAGAG--ACGGTGGGTAAAGCATGTGGGCACGACGACGAGG 3060
QY 2123 aaacaacagaaacacaaggaagcgtgtgtggtggtggtggtggtggtggtggtggtggtg 2181
Db 3059 AACACAACAG-----GGGAGGTGGCTGTGGCCGGGCTGAGCTGTGCTGGGGTTTGTG 3007
```

Db 150367 CGAGCCCCAGAGCTAAAGCACCAGAGAGAAGTCAAAATGCTTCCATTTCAGCATGACCTAA 150308
QY 2480 ctttcttagtgtgcacgagccccaccacctcctgcagtlaccacacacatcaacctgtcgt 2539
Db 150307 C--TTCTAGTGTGCCACGGCCCCCACCACCTTCCGTGAGTGCCCCACACATCACCTGCTT 150250
QY 2540 tctcttccacagtgatcgtatcttcttagttcattatcttctttagatgat----- 2594
Db 150249 TCTCTTCCACAGTGAATGTATTTCTTAGTTTCATTAATTTCTTTTATGTGTATGACAC 150190
QY 2595 ----- 2594
Db 150189 TATATAAATTTTCCCTCTCCCTGCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCT 150130
QY 2595 ----- 2594
Db 150129 TTCTTTCTTTCAAGACAAGTCTCACTCTGTCACCCAGGCTGAGTGCACGGCAGAGTC 150070
QY 2595 ----- 2594
Db 150069 TTGGCTCCCTGCACACTTCTGCCCTCCAGGTTCAACAATTGTCTGCCCTCAGCCTCCCGAG 150010
QY 2595 ----- 2594
Db 150009 TAGCTGGGCCACACACACCAGCTAGTTTGTACTTTTATAGAGAGGGGTTTCACCAT 149950
QY 2595 ----- 2594
Db 149949 CTGGGCCAGGCTGTGTCTTGAACCTCTGACCTGTGATCCACCCGCTCGGCTCCCAAG 149890
QY 2595 -----gacctatataaatcttcattga 2619
Db 149889 TGCTGGGATTACTGGCATGAGCCACCGTGCCCGACACACTATATAAAATTTTCATTGA 149830
QY 2620 gaatttctcaattgtatcttagttaatagcacagtttgaacctgtctgagactgactt 2679
Db 149829 GAGTTTCTCACTGTATCTAATATATATCGCACAGTTTGGAACCTTTCTGAGACTGACTT 149770
QY 2680 tatcaataatctaaccgacaaagatcatatccatgtgtatgtgttagaacatttattt 2739
Db 149769 TATCAATATATTAACCGACCAAGATCATATCCAGCTGATGTAGTATGACATTTTATTT 149710
QY 2740 cattgactaaccagagacagtttcagtgatgacaaattgtgtccctcgtgttcagctgaa 2799
Db 149709 CATGACTAACCCAGACCACTTTCAGTGATGCAAAATTACGTGCCCTATGTGTCAGCTGAA 149650
QY 2800 acagtcctgacttcaaaaaacctgaataagtcctccacacagttgtataaatggacaat 2859
Db 149649 ACAGTCCCTGACTTTCAAAAACCTTGAATTAAGTCTCCACAGTTGTATAAATTGGACAAT 149590
QY 2860 ttaaggaattttaaactttagatgatcatcttggttccatttatttatttatttattt 2919
Db 149589 TTAGGAATTTTAAACTTTAGATGATCATTTGGCTCCATTTTATTTTATTTTATTTTG 149530
QY 2920 ttaatgcaaacagagactt-----aatgaacttggatctctgtttttaaagattatta 2975
Db 149529 TTAATGCAACAGGACTTTAAACAATGAACCTTTCATTTTAAATATTTCTTTAAGA 149470
QY 2976 acattgtgtatcatatagatgctccttgagacttagcttccactacacagatat 3035
Db 149469 AAATCATGTGTCTATACATACGGGCTCTTGAGGACTTACCTTTCACTACACTACAGGATAT 149410
QY 3036 gatctccatgtagtcataataaacctgcagagtgatttccagagtgctcgatactgtta 3095
Db 149409 GATCTCCACATAGTTTACATATAAACTGCAGACTGATTTTCCAGAGTGTCTGATACTATTA 149350
QY 3096 attacatctcattagtgctgaaagaatgacctacglttctgtatatacagctgtgtgct 3155
Db 149349 ATTACATCTCGATTAGGCTGAAAGATATGACCTATGTTTCTTTATACAGTGTGTGCT 149290
QY 3156 tttagatgtgtgtactgtacacagaagtgtgtcactgaggc 3198
Db 149289 TTTGATGTTGTGTACTATACACAGAAGTGTGTGCACTGGGCC 149247

RESULT 12
AC068845/c
LOCUS
DEFINITION
Homo sapiens chromosome 19 clone RP11-73711 map 19, WORKING DRAFT
SEQUENCE, 27 unordered pieces.
AC068845
AC068845.3 GI:9966312
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 194407)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 19, clone RP11-73711
Unpublished
2 (bases 1 to 194407)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczký,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneses,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
COMMENT
Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:8247880.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7549
Center clone name: 737_I_1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181875 bases at least Q40
Consensus quality: 187842 bases at least Q30
Consensus quality: 190220 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 191807; sum-of-ctnigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

Db 1366 GGGCACCCTGCCACAGCCGAGACCATTTTAAAGAGTAACATCTCTTCAGACATGTACC 1425
QY 1696 ccatgacctctcacgagacccttgagcaactgactatcttcagagtgccaggatt 1755
Db 1426 CCATGACCTCGCACAGACCCTTGACCACTGTACTACTCTTCCAGAGCCAGGATT 1485
QY 1756 ccaggtgaatacaaaagacttccccaacaaacagaagaattgtatctctatcaa 1815
Db 1486 CCAGGTGATATACAAAGACTTCCCAAGAACACAGACAGAGTGTATCTCTCATCAA 1545
QY 1816 ttgctccttcagccacctctgatacagcatgtgtatcgcaagatgtgagtcctgcca 1875
Db 1546 CTGCTCTTCACAGCCACCTCTCGTCAGCCATGGCATCGGCAAGATGTGAGTCTCTGCA 1605
QY 1876 tgatatgctgcgtgcaacatcttaaaagltgctgtcgtgagtggaaccaagaatcaga 1935
Db 1606 TGATATGGCTGCACCTGACACATCTTAAAGCTGCTGTGAGTTGGACCAACAGAGACAGA 1665
QY 1936 gatgccaagaacaggaacagcaatgtctgtgtgagagtgctgaaccttctcg 1995
Db 1666 GATGCCAAGACAGAAACGAGCAGTTTTCAGCGTGGGAGGTGCTGAACCTATCTCG 1725
QY 1996 ccatgaaccattataaatacccaatatatactgaanaactga-aactgcttgaanaa 2054
Db 1726 CCATGAACCATTTATAAA-CCCAACATATATACCGAAATACTGAGAAGCTTTGAA 1784
QY 2055 ttgtgaattctctgatacctccagtggtgagcgagaga 2089
Db 1785 TTGGAAATTTCTGATACTCCAGTGGGCCAAGACA 1819

RESULT 11
AC011492/c AC011492 157803 bp DNA HTG 04-MAY-2000
DEFINITION Homo sapiens chromosome 19 clone CTB-187L3, WORKING DRAFT SEQUENCE,
8 unordered pieces.
AC011492
AC011492.5 GI:7690131
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 157803)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157803)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On May 4, 2000 this sequence version replaced gi:7458736.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 136279, BC101302
Center clone name: CIT978SKB_187L3

Summary Statistics
Consensus quality: 152548 bases at least Q40
Consensus quality: 153824 bases at least Q30
Consensus quality: 154424 bases at least Q20
Estimated insert size: 156750; agarose-fp estimation
Estimated insert size: 157103; sum-of-contigs estimation
Quality coverage: 9.52 in Q20 bases; agarose-fp estimation
Quality coverage: 9.5 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 2375 2374: contig of 2374 bp in length
* 2475 2474: gap of unknown length
* 4521 4521: contig of 2047 bp in length
* 4522 4621: gap of unknown length
* 4622 7566: contig of 2945 bp in length
* 7567 7666: gap of unknown length
* 7667 12083: contig of 4417 bp in length
* 12084 12183: gap of unknown length
* 12184 18288: contig of 6105 bp in length
* 18289 18388: gap of unknown length
* 18389 63650: contig of 45262 bp in length
* 63651 63750: gap of unknown length
* 63751 111373: contig of 47623 bp in length
* 111374 111473: gap of unknown length
* 111474 157803: contig of 46330 bp in length.
FEATURES
Location/Qualifiers
Source 1. 157803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTB-187L3"
/clone_lib="CalTech human BAC library B"
BASE COUNT 40983 a 38390 c 39640 g 38079 t 711 others
ORIGIN

Query Match 21.1%; Score 705; DB 35; Length 157803;
Best Local Similarity 72.5%; Pred. No. 1.8e-155;
Matches 1206; Conservative 0; Mismatches 90; Indels 367; Gaps 8;

QY 1883 gctgcgtgaacatcttaagttgctgtctgagtggaaccaaaagtlacagagatgcca 1942
Db 150889 GCTGCACCTGAACACCTTAAAGTTGCTGTGAGTTGGACCAACAAGTACAGAGATGCCA 150830
QY 1943 agaacaagaaacgacccaatgtctgtgtgtgagagtgctgaaccttltctgacctgaa 2002
Db 150829 AGAACAGGAATGATCAATGTCTGTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAA 150770
QY 2003 ccatataaaatcccaacatatatactgaaataactgaactgcttgaanaatttgaat 2062
Db 150769 CCATTATAAATPCCCAACATATACACCTGAGAAATATTAAACTGCTTTGAAATTTGAAAT 150710
QY 2063 ttctgatacctcagtggtggcagagagacagtggtgtaagatgtggtgacagcagcagg 2122
Db 150709 TTCTGATACCTCTAGTGGCCGAGAG--ACGGTGGGTAAGAGATGTGGGACAGCAGCGG 150652
QY 2123 aagacaacagaacacacaaggaagcgtgtgccc-ggctgagatgtgtggtgttttg 2181
Db 150651 AACACAACAG-----GGAGGTGGCTGTGGCCGGGCTGGAAGTGTGCTGGGGTTGTG 150599
QY 2182 tgatggcactcgtgtgacctggtgcgttccttaacgaatagcagctgcctgtggtgaagaag 2241
Db 150598 TGACGGCCACTGCTGACCTGCTGCTGTCTTACGCAATACACAGCTGCTTTGGGGAAGAG 150539
QY 2242 ggctgccagcagcagctgtgtctcccggtgacacagcagatccacacccctgtgacctcg 2301
Db 150538 GGCTGCCACAGCCAGCTGGCTCTCCTGGGGCACACAGCAGATCCACGCCCTGGGCACCTTCC 150479
QY 2302 tgttggtcttttttccctgtg--tgaagaagaacgacagcccttctcaagc 2359
Db 150478 TGTTTGCTCTTTTTCCTCGGTGAAGAAGAAACTGAAGAGATGACCCCTCTCTCAAGC 150419
QY 2360 tggctcactcagacacatttggtgacaacacctgtgacagccatgccaagagagagccttga 2419
Db 150418 TGGCTTAG-----TTGGACAAACCTCGACAGCCACACAGAGAGAGGCTTCAA 150368
QY 2420 ccggcccagagctaaagcaccagagaaatcaaatgtcttctactcagcgtgacccaa 2479
|| |||||

Db 1878 TCCAGTGGCCAGACA 1894

RESULT 10

AF227200 3059 bp mRNA ROD 09-AUG-2000

LOCUS Rattus norvegicus staufen isoform Stau+I6 (Stau) mRNA, complete

DEFINITION cds, alternatively spliced.

AF227200

AF227200.1 GI:9754865

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 3059)

AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desgroselliers, L., Kuhl, D., Richter, D. and Kindler, S.

TITLE Two rat brain staufen isoforms differentially bind RNA

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3059)

AUTHORS Monshausen, M. and Kindler, S.

TITLE Direct Submission

JOURNAL Submitted (20-JAN-2000) University of Hamburg, Institute for Cell Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg 20246, Germany

FEATURES

source location/Qualifiers

1. .3059

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/issue_type="brain"

/dev_stage="adult"

1. .3059

/gene="Stau"

227. .1714

/gene="Stau"

/note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau+I6"

/codon_start=1

/product="staufen isoform Stau+I6"

/protein_id="AAF98119.1"

/db_xref="GI:9754865"

/translation="MYKPYVDHPSRMOSTYSTYGMKRGAVPPRYFPVPVPLLQVELS VGGQFNGKGMRPVKHDATARALRTIQSEPLPERLEVNGRESEENLNKSEISQVF EIALKRNLPVNFESEFPLTYVARESGPPHMKNFVTRVSGEFGEGESKTSKRNAA RAVLEQLRLPLPAVERVKPRTKKKSQPTCKLQTPADYGOGMNPISRLAQIQAKKE KEPEYMLTERGLPRRREFVMOVKVGHHTAEGAGTNKRVAKRNAENMLEIGFKVPQ AOPAKPALKSEKTPVKKPGDKRKVTFEFSPGDENGTSNKDEFRMPYLSHQQLPAG ILPWPEVAQAVGVSQGHHTDFARAAPNPAKATVPMIARELLYGTSPTAETILKS NISGHPHGPRTRPSEQLYLSRAQGFQVEYKDFPKNNKNECVSLINCSSQPPLYVSH GIGKDVESCHDMAALNILKLSELDOOSTEMPRTGNGPVSACGRC"

polyA_signal 3043. .3048

/gene="Stau"

BASE COUNT 817 a 778 c 749 g 713 t 2 others

ORIGIN

Query Match 37.4%; Score 1253.8; DB 74; Length 3059;

Best Local Similarity 86.3%; Pred. No. 8.2e-285;

Matches 1428; Conservative 0; Mismatches 207; Indels 20; Gaps 3;

QY 454 agaaagcataaccctactgtagaactaatgcaactgtgcatgaaacttgyaaaaaac 513

||||| ||| | |||| | ||| | ||||| ||||| ||||| ||||| |||||

Db 166 AGAAGTATAGCTTCTACCATTGAGCTCAATGCACCTGTGCGTGAAACTGGAAAGAAACC 225

QY 514 aatgtataagcctgttgacccttactctcgtgagtcagtcacactataactacaacatgag 573

||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 226 AATGTATAAGCCCGTGAGACCCCCACTCTCGGATGCAGTCCACCTACAGCTATGGCATGCG 285

QY 574 aggagtgcttatcccccagagtlacttaccacatttcagttccaccttacttatca 633

Db 286 TGGAGGTGCTTATCCCCCAGATACTTTTACCACATTTCCAGTTCCACCCGTTACTCTACCA 345

QY 634 agtgyaacttctctgtggygagcagcaatttaatgcaaaagaaagacagcagctgc 693

||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 346 AGTTGAGCTCTCCGTGGGAGACAGCAGTTTAAATGGGAAGGAAGATGAGACCACCTGT 405

QY 694 gaacacgatgctgctgctccaaagcgttgagatccctgcagaatgagccctgcagagag 753

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 406 GAACATGATGCCACTGCCCCAGGCCCTGAGGACTCTGCAGAGTGAGGCCCTTACCAGAGAG 465

QY 754 gctggaagtgtaatggaagaaatccgaagaagaaatctcaataaatctgaataagtca 813

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 466 GCTAGAGGTAATGGAAGAAATCAGAAGAAGAAACCTTAATAATCAGAATAAAGCCA 525

QY 814 agtgttgagattgcaactaaacggaacttgctgtgaattcg----- 857

||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||

Db 526 AGTGTGTGAGATTGCGCTGAAGCGGAATTTGCCCTGTGAATTTTGAGTCTTCCCTCTGAC 585

QY 858 --agtgccccggagagtggtgccccacccacatgaagaacttgtgaccaagttcgt 915

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 586 ACAGGTGCCCCGGAGAGTGGCCCCGCCACACATGAAGAACTTTGTGACCGAGGTTTCAGT 645

QY 916 tgyggaagttgttgyggygaagtgtaaggygaaagcaagaagatttcaaaagaaatgcygc 975

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 646 TGGGGAGTTTGTAGGGGAGAGGAAGGAAAGCAAGATCTCCAAGAGAATGCGGC 705

QY 976 catagctgttcttgagagcgtgaagaagtlaccgccccctgctcgtcagltgaacgagtaaa 1035

| || || || || || || || || || || || || || || || || || || || || || ||

Db 706 AAGGCCCTGCTGTGAGACGCTAAGGAGGCTGCCACTCTACCTGCTGTGTGAGCGAGTGAA 765

QY 1036 gcttagaatcaaaagaagaacaaaccatagtcacagccacagacaagccagaatatg 1095

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 766 GCCCAGATCAAGAGAAGAAAGTACAGCCACCTGCAAGCTACAGACTGCCCCGACTATAGG 825

QY 1096 ccaggggatcaatccgatagccgactggtgccccagatccagcagcaaaaaagagaagga 1155

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 826 CCAAGGATGATTCCTATTTAGCAGACTGGCACAGATCCAGCAGCAAAAGAGAAGGA 885

QY 1156 gccagagtacaagctcctcaacagagcagcgtcccgcgccgagaggttgtgatga 1215

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 886 GCCGAGTACATGCTCCTTACAGAACGAGGTCTTCACGCTCGCAGGAGATTGTGATGCA 945

QY 1216 ggtgaaagttggaaccacactgcagaaggaacggygcaaccaagaagaagttgccaagcg 1275

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 946 GGTGAAGGTTGGGCATCACACCGCAGAAGGAGCAGTACCAATTAAGAAGGTGGCCAAAGCG 1005

QY 1276 caatgcagccgagaacatgctgagatccttggttcaaaagtcgccgagcgagccac 1335

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1006 CAATGCTGCTGAGAACATGCTGAGATCCTGGGGTTCAAAGTCCCCAGCGCAGCCTGC 1065

QY 1336 caaaccgcactcaagtcaagaggaagaagacaccataaagaacccagggatggaagaaa 1395

||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1066 CAAGCCAGCACTCAAAATCAGAAGAGAAGAGACTCCAGTAAAGAAACCAAGCATGGAAGAA 1125

QY 1396 agraacctttttgaaacctggtctggygatgaaaaatggyactagtaataaagagatga 1455

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1126 AGTACATTTTGTGAACCTAGCCCTGGGATGAATAATGGAACCACTAATAAGAGATGA 1185

QY 1456 gtccagatgccttatcttaagtcatcagcagcgtcgtcgtgaaattcttcccatggtgc 1515

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1186 GTTCAGGATGCCCTATCTTAAGCCATCAGCAGCTGCCAGCCGGGATTTCTCCCATGTGCC 1245

QY 1516 cgaagtcgcccagagcgtgtgagagltagtcaaagacatcacaccaagaagatttaccagggc 1575

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1246 AGAGGTTGCCCAGGCTGTGCGGGTTAGTCAAGGACACCATACCAAAAGATTTCGCCAGGGC 1305

QY 1576 agtccgaatccctgcgaagyccaagytlaactgcatgataagccgagagtggtglatag 1635

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1306 AGCTCCAAATCCTGCCAAGGCCACGCTAACTGCCATGATAGCCCCGAGAAATTGTGTACGG 1365

QY 1636 gggcacctcgccacagccgagacattttaaagataaacatctcttcaaggccaagtlacc 1695

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

/product="staufen isoform Stau-16"
/protein_id="AAF98177.1"
/db_xref="GI:9755163"
/translation="MYKVPDPHSMQSTYSYGMRGAGYPPRYFPYPPVPLLQYVELS
VGGQFNGKGMKRPVKHDATARALRLQSEPLPERLEVNGRESEENLNKSEISQYF
EIALKRLNLPVNEFVARESPPHMKNFVTRVSVGEFVGEQSKSKISKNNARAVLEQ
LRRLPLPAVERVKPRIKKSQPTCKLQIOTAPDYOGMNPISRLAQIQAKKEKEPEYM
LTERGLPRRREVMQVKVGHHTGAGTNKKVAKRNAENMLEILGFKVPOAPKPKP
ALKSEKTPVKKPGDGKVTFFEPSGDENGTSNKEDFRMPYLHQOLPACILPMVP
EVAQAVGSQGHHTKDFARAAPNPAKATVTAMIARELYGTSPTAETILKSNISSGH
VPHGPRTRPSEQLYLSRAQGFQVEYKDFPKNNKNECVSLINCSQPLVSHGIGKDY
ESCHDMAALNLLKLLSELDOQSTEMPRTGNGPVSACGRC"
polyA_signal 3025..3030
/gene="Stau"

BASE COUNT 814 a 771 c 747 g 707 t 2 others
ORIGIN

Query Match 38.3%; Score 1281.8; DB 74; Length 3041;
Best Local Similarity 87.2%; Pred. No. 2.1e-291;
Matches 1428; Conservative 0; Mismatches 207; Indels 2; Gaps 2;

QY 454 aaaaaagcataaaccctactgttagaactaaatgcactgtgcatgaaacttggaaaaaacc 513
Db 166 AGAAAGTATAGCTTCTACCATTTGAGCTCAATGCTGCTGCTGAAACTGGAAAGAAACC 225
QY 514 aatgtataagcctgttaacccttactctcgtatgcagtcacactataactacaacatgag 573
Db 226 AATGTATAGCCCGTGAGCCCACTCTCGATGAGTCCACCTACAGCTATGGCATGGC 285
QY 574 aggaagtgctatacccccaggtactttaccattccagttccaccttacttatca 633
Db 286 TGGAGCTGCGCTATCCCCAGATACTTTACCCATTCCAGTTCCACCGTTACTCTACCA 345
QY 634 agtgaacttctgttgagagagacagcaatttaatgycaaaagaaagacagagcgtgc 693
Db 346 AGTTGAGCTCTCCGTGGGAGACAGCAGTTTAAITGGGAAGGGAAGATGACACCCTGT 405
QY 694 gaaacacgactgctgtgcccaaaagcgttgagatccctgcagaaatgagccctgccagagag 753
Db 406 GAAACATGATGCGCACTGCGCCGAGCCCTGAGGACTCTGCAGAGTGAGCCCTTACCAGAGAG 465
QY 754 gctgaggttgaatggaagaaatccgaagaagaaatctcaataaatctgaataatga 813
Db 466 GCTAGAGGTAATGGAAGAAATCAGAAGAAAACTTAAATCAAGAAATAAGCCA 525
QY 814 agtgtttgagatgtcacitaaacggaactgtcctgtaatttcgagtggtgcccggagag 873
Db 526 AGTGTTTGAGATTGCGGTGAAGCGGAATTGCTGTGAATTTTGAGGTGGCCCCGGAGAG 585
QY 874 tggccccaccacatgaagaactttgtgaccaaagtttcggttgaggagtttgtgggga 933
Db 586 TGGCCCCGCCACATGAAGAATTGTGACACAGGTTTCAGTTGGGGAGTTTGTAGGGGA 645
QY 934 aggtgaagggaaaagcaagaagatttcaaaagaaaatgccgcataagctgttcttgagga 993
Db 646 AGGAGGAAGGGAAAAAGCAAGATCTCCAAGAAATGCGGCAAGGGCCGTGCTGAGCA 705
QY 994 gctgaagaagtlaccgcccctgcctgcagtltaacgagtaaaagcctagaaatcaaaaaga 1053
Db 706 GCTAAGGAGGCTGCCACTCTACCTGTGTGAGCGAGTGAAAGCCAGAAATCAAGAAAGAA 765
QY 1054 aacaaaaccatagtlcaagccacagacaagcccaagaatatgcccagggatcaatccgat 1113
Db 766 AAGTCAGCCCACTGCAAGCTTACAGACTGCCCCGCACTATGGCCAAGGATGAATCCTAT 825
QY 1114 tagccgactggcccagatccagcagcaaaaagagaagagccagagtatacagctcct 1173
Db 826 TAGCAGACTGGCACAGATCCAGCAGGCAAAAAGAGAAGAGCCGAGCTACATGCTCCT 885
QY 1174 cacagagcagagcctcccgcgccgcagggagtttgtgatagcaggtgaaagttgaaacca 1233
Db 886 TACAGAACGAGGTTCTCCACGTCGACAGGAGTTTGTGATGACAGGTGAAGGTTGGGCATCA 945

QY 1234 cactgcagaaaggaaacggcaccaacaagaaggttgcccaagcgcaatgacgcgaagaacat 1293
Db 946 CACCGCAGAAAGAGCAGAGTACCAATTAAGAAGGTGGCCAAAGCCCAATGCTGTGAGAACAT 1005
QY 1294 gctgagatccttggtttcaaaagtcgccgagcgcgagcccaaccaaacccgcactcaagtc 1353
Db 1006 GCTGAGATCCTGGGGTTCAAGTCCCCAGGCGCAGCCTGCAAGCCAGCACTCAAAATC 1065
QY 1354 agagagaagacacccataaagaaacccagggagtgaagaaaaagtaacctttttgaacc 1413
Db 1066 AGAAGAGAAGACTCCAGTAAGAAGAACCCAGAGATGGAAGAAAGTAACATTTTGAACC 1125
QY 1414 tggctctgggagtgaaaaatgaggactagtaataaagagatgagttcaggaatgcttatct 1473
Db 1126 TAGCCCTGGGGATGAATAATGGAACCAAGTAATAAGAGATGAGTTCAGGATGCCCTATCT 1185
QY 1474 aagtcacagcagctgcctgtgtaattcttcccatggtgccgaggtgcgccagctgt 1533
Db 1186 TAGCCATCAGCAGCTGCCAGCGGGATTTCTCCCATGTGTGCCAGAGGTGCCAGGCTGT 1245
QY 1534 aggagttagtlcaagacatcacaccaagaatttaccagggcagctccgaatcctgcca 1593
Db 1246 CGGGGTTAGTCAAGGACACCACTACCAAGATTTGCCAGGGCAGCTCCAATTCCTGCCAA 1305
QY 1594 ggcceacgtaactgccaatgtagcccgagagttgtgtatggyggcaccctgcgccacagc 1653
Db 1306 GGCCACGGTAACGTGCCATGATAGCCCGAGAATTGTGTACGGGGGCACCTGCCACAGC 1365
QY 1654 cgaagaccatttaagaataacatctctcagggccaagtaaccatgagctctcaagag 1713
Db 1366 CGAGACCATTTTAAAGAGTAAACATCTCTTCAGAGACATGTACCCCATGAGCTTCGACAAG 1425
QY 1714 accctctgagcaactgtagatacttccagagtlccaggagatccaggttgaatacaaaaga 1773
Db 1426 ACCCTTGAGCAACTGTACTACCTTCCAGAGGCCAGGGATTCAGGTTGAATACAAAGA 1485
QY 1774 ctccccaaaacaacaagaacgaatttgtatatcttatacaattgctcctcagccacc 1833
Db 1486 CTTCCCAAGAACAACAAGAGAGTGTATCTTCATCAACTGCTTTCACAGCCACC 1545
QY 1834 tctgatcagccatggtatcggaagatgtgtgagtcctgcatgatatgctgcgtgaa 1893
Db 1546 TCTGCTCAGCCATGGCATCGGCAAGGATGTGGAGTCTTCATGATATGGCTGCACCTGAA 1605
QY 1894 catcttaagttgctgtctgaagttggaaccaacaagttacagagatgccaagaacagga 1953
Db 1606 CATCTTAAAGCTGCTGTCTGAGTTGGACCAACAGACACAGAGATGCCAAGCAGAGAAA 1665
QY 1954 cggaccaatgtctgtgtgaggaggtgtgtaacctttctgycatgaaccattataaa 2013
Db 1666 CGGACCAATTTCAGCGGTGCGGAGGTGCTGMAACTTATCTGGCCATGAACATTATAAA 1725
QY 2014 tcccaacatatatactgaaataactga-aactgcttgaanaatttgaattctgatacc 2072
Db 1726 -CCCAACATATATATACCGAAATACTGAGAACTGCTTGAATAATTCTGATAAC 1784
QY 2073 tccagtgggccgagaga 2089
Db 1785 TCCAGTGGGCCAAGACA 1801

RESULT 9
AF061942 2860 bp mRNA ROD 04-MAR-1999
LOCUS
DEFINITION Mus musculus RNA-binding protein stauflen mRNA, complete cds.
ACCESSION AF061942
VERSION AF061942.1 GI:4335944
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

/clone="RP3-470L14"
/clone_1ib="RPCT-3"
misc_feature 3061..3077
/note="Weak data"
misc_feature 95583..95627
/note="Single clone region. pUC subcloned to give 12x coverage. Assembly confirmed by restriction digest data."
BASE COUNT 33974 a 25115 c 24233 g 30514 t
ORIGIN

Query Match 40.5%; Score 1357.4; DB 34; Length 113836;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1969 gtgtggaggtgtctaaaccttctctgcccataaaccattataaaatcccaacatatatc 2028
Db 34965 GTGTGGAGGTGCTGAACCTTTCTGTGGCCATGAACCATTTATAAAATCCCAACATATATAC 34906
QY 2029 tgaataactgaactccttgaataatttgaaattctgataccctcagtgggccgagag 2088
Db 34905 TGAATAACTGAACCTGCTTGAATAATTGGAATTTCTGATACCTCCAGTGGGCGGAGAG 34846
QY 2089 acaagtggtgaagaagtgtggcagcaggggaagacaacagaacaacaagagcg 2148
Db 34845 ACACGCTGGGTAAGAGATGTGGGACAGCAGCGAAGACAACAGAAACACAAGAGGCGG 34786
QY 2149 ctgtggcc-ggctggaactgtctgtgggttctgtatgagccactcgtgacctgagcg 2207
Db 34785 CTGTGGCCGGCTGAGCTGTGCTGGGGTTGTGTGTGATGCCACATCGGTGACCTGGCGGT 34726
QY 2208 cccatgcataagcagctcctgtgtgggaagaaggctgcccagcagctgttctccg 2267
Db 34725 CCTACGCATAGCAGCTGCTGTGGGGAAGAGGGCTGCCACGCCAGCTGTCTCCCG 34666
QY 2268 ggaacaccagcagatccacacccctgggcaacctcgtgttcttccctgtgt 2327
Db 34665 GGACACCAAGCAGATCCACACCCCTGGGACCTCCGTGTTGTTTCCCTGTGT 34606
QY 2328 gaaagaagaagaacgagcagccctctcaagctggctcactcagacacattggacaac 2387
Db 34605 GAAAGAAGAAGCGGACAGACCCCTTCTCAAGCTGGCTCAGTCAACATTTGGACAAC 34546
QY 2388 cctggacagccatgcccagaagagcgcttggaccggccccagagctaaagcaccagaga 2447
Db 34545 CCTGGACAGCCATGCGCAGAGAGGCGCTTTGACCGGCCCCAGAGCTAAAGACACAGAGA 34486
QY 2448 aatcaaatgtcttctactcagcgtgacccaactttctagtgtgccaagggcccaaccac 2507
Db 34485 AATCAATGTCTTCCTACTCAGCGTGACCCCACTTTTCTAGTGTGCCACGGCCCCACAC 34426
QY 2508 ctctctgagtaaccacacacatcaccactgcttctcttccaacagtgatctgtattcta 2567
Db 34425 CTCTGCAGTACCCACACCATCACCACTGCTTCTCTCCAACAGTGAATCTGTATCTTA 34366
QY 2568 gttcattatttcttcttgattgatatgacacataataaatttcattgagaattct 2627
Db 34365 GTTTCATTATTTTCTTTGATTGATATGACACTATATAAAATTTTCATTGAGAAATTTCT 34306
QY 2628 caatgtlactagtltaaatagcacagtttggaaactgtctgagactgacttacaata 2687
Db 34305 CAATGTATCTAGTTAAATAGCACAGTTTGAACCTGTCTGAGACTGACTTTATCAATA 34246
QY 2688 atctaaccgacaaagatcatatcatatgtgtatgtgttagacattttattcatgtact 2747
Db 34245 ATCTAACCGACAAAGATCATATCCATGTGTATGTGTTAGACATTTTATTTCAATTGACT 34186
QY 2748 aaccagagacagtttcagtgatgcaaatgtgtgcccctctgttccagctgaacagtcct 2807
Db 34185 AACCCAGACAGATTTCAGTGAATGCAAAATGTGTGCCCTCTGTGTCAGCTGAACAGTCTCT 34126
QY 2808 ggactttcaaaaaccttgataagtctccacagttgtataaatggacaatttagaat 2867
|||||

Db 34125 GGACTTTCAAAAACCTTGAATAAGTCTCCACAGTTGTATTAATTGACAAATTAGGAAT 34066
QY 2868 tttaactttagatgatcatatttggttccattttatttcaattttatttctgttaatgca 2927
Db 34065 TTTAACCTTAGATGATCATTTGGTTCATTTTATTTCATTATTGTTGTTAAATGCA 34006
QY 2928 aacaggaacttaatagaactttagatctctgtttttaaagaattataaaacacattgtgtatc 2987
Db 34005 AACAGACTTAATGAACCTTGTATCTGTCTTTTAAAGATTATTAATAAACATTGTGTATC 33946
QY 2988 tatacatatgctctttaggactttagactttagcttcaactacacagatatgatctccatgta 3047
Db 33945 TATACATATAGGCTCTTGAGGACTTAGCTTTCACTACACTACAGATATGATCTCCATGTA 33886
QY 3048 gtccatataaaccctgcagagtgatatttccagagtgctcgatactgttlaatacatcca 3107
Db 33885 GTCCATATAAACCTGCAGAGTGATTTTCCAGAGTGCTCGATCTGTTAATTACATTCOA 33826
QY 3108 ttagggctgaagaatgacactacgttctctgtatacagctgtgttctttagtgtgtg 3167
Db 33825 TTAGGGCTGAAGAAGATGACCTACGTTTCGTATACAGCTGTGTGCTTTGATGTGTG 33766
QY 3168 ttactgtacacagaagtgtgtgcacttagagctctgcgtgtgtccgtatggaacactgg 3227
Db 33765 TTACTGTACACAGAGAGTGTGTGCACGTAGGCTCTGCCGTGTGTCGGTATGGAAGCCCTGG 33706
QY 3228 tagccctgcgagttgaagtactgtcttccatcattgtttaaagctggaattttctcccat 3287
Db 33705 TAGCCCTCGAGTTAAGTACTGCTTCCATTCAATTGTTTACCGCTGGAATTTTCTCCCAT 33646
QY 3288 ggaatgtgaataaacttaagtgttctgcatcaataaataatgtaataactaaa 3338
Db 33645 GGAATGTAAATAAAGTTAAGTGTGTGTCATCAATAAATGTTAACTACTAAA 33595

RESULT 8
AF290989
LOCUS AF290989 3041 bp mRNA ROD 09-AUG-2000
DEFINITION Rattus norvegicus staufen isoform Stau-16 (Stau) mRNA, complete cds, alternatively spliced.
ACCESSION AF290989
VERSION AF290989.1 GI:9755162
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3041)
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., DesGroselliers, L., Kuhl, D., Richter, D. and Kindler, S.
TITLE Two rat brain staufen isoforms differentially bind RNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3041)
AUTHORS Monshausen, M. and Kindler, S.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) Institute for Cell Biochemistry and Clinical Neurobiology, University of Hamburg, University Hospital Eppendorf, Suderfeldstr. 24, Hamburg D-22529, Germany
FEATURES
source
1..3041
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1..3041
/gene="Stau"
227..1696
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau-16"
/codon_start=1

Db 1559 CAACATATATACTGAAATACTGAAACTGCTTTGAAAAATTGCAATTCTGATACCTCCA 1618
QY 2077 gtgggcccagagacacacggtggtgtaagatgtggtgcagcagcaagggaagacacagaaac 2136
Db 1619 GTGGCCGAGAGACACGCTGGGTAAGGATGTGGGCAGCAGCAGGGAAGACACAGACAAAC 1678
QY 2137 acaagagagcggtctgtggcc-ggcttggactgtgtctgggttttgttgtatggtccactcgg 2195
Db 1679 ACAAGGAGGCGGCTGTGGCCGGGCTGGACTGTGCGGGGGTTGTGTGATGGCCACTCGG 1738
QY 2196 tgacctggcggtccctacgcacatagcagctgcctgtgggaaagaaggctgccccag 2255
Db 1739 TGACCTGGCGGTCCCTACGCAATAGCAGCTGCCCTGTGGGGAAGAGGGGCTGCCCAAG 1798
QY 2256 ctggttctccgggacacacagcagatccacacccctgggcacccctcglttgtctttt 2315
Db 1799 CTGTTCTCCCGGACACACAGCAGATCCACCCTGGGCACCTCCGTGTTGTCTTTT 1858
QY 2316 ttccctgtgtgaaagaagaacggcacgacccctctcaagtgtgtcactcagacac 2375
Db 1859 TTTCCCTGTGTGAAGAAGAACGGCAGACGCCCTTCTCAAGCTGGCTCAGTACAGAC 1918
QY 2376 attggacaaacccctgagcagccatgacagagagagagcctttgacggccccagaagctaa 2435
Db 1919 ATTGGGACAAA-CCTGGACAGCCCATGCCAGAGAGAGG-CTTTGACCGGGCCCCAGACTAA 1976
QY 2436 aagcacagagaaaaatcaaatgtctctactcagcgtgaccccaactttctagtgtgcca 2495
Db 1977 AAGCACACAGAAAAATCAAAATGTTCTTACTCAGCGTGACCCCACTTTCTAGTGCCA 2036
QY 2496 cggccccacacccctcctgcagtaaccacacacatcaccaactgtctctccaacgtga 2555
Db 2037 CGGCCCCACACCTCTCGACGTACCCACACCATCACCACTGCTTCTCTCCACAGTGA 2096
QY 2556 tctgtattcttagttcatatttcttttgtattgatgacactatataaaatttcat 2615
Db 2097 TCTGTATCTTAGTTTCATTATTTCTTTTGATGTGATATGACACTATATAAATTTTCAT 2156
QY 2616 ttgagaattctcaatttgtatctagttaaatagcacagtttgaaacttgtctagactg 2675
Db 2157 TTGAGAAATTTCTCAATGTGTATCTTAATAGCACAGTTTGAAACTGTCTGAGACTG 2216
QY 2676 acctatcaataatctaaccgcagaaagatcatatccatgtgtatgtgttagacatttt 2735
Db 2217 ACTTTATCAATAATCTAACCCGACAAAGATCATATCCATGTGTATGTGTAGACATTTT 2276
QY 2736 attcatgactaacccagagacagtttcagtgatgacaaattgtgcccctcgttcagc 2795
Db 2277 ATTTCATTCATAACCCAGGACGTTTCAGTGATGCAAAATGTGTGCCCTCTGTCAGC 2336
QY 2796 tgaacagctcctgacttcaaaaaaaccttgaataagtcctccacagttgtataaatgga 2855
Db 2337 TGAACAGCTCCTGGACTTCAAAAAACCTGAATAAGTCTCCACAGTGTATATAATTGGA 2396
QY 2856 caattaggaattttaaacttagatgatcatattgttccatttttatttatttatt 2915
Db 2397 CAATTTAGGAATTTTAAACTTTAGATGATCATTTGGTTCCATTTTATTTTATTTAT 2456
QY 2916 ttgttaatgcaaacagagacttaaatgaaacttgatctcgtltttaaagatataaaaa 2975
Db 2457 TTTGTTAATGCAAAACAGGACTTAAAGAACTTTGATCTCTGTTTAAAGATTTATAAAA 2516
QY 2976 acattgctatctatatacatattgctcttgagacttagcttcaactacactacagatat 3035
Db 2517 ACATTTGTATCTATATACATATGGCTCTTGAGGACTTAGCTTTACACTACAGAGATAT 2576
QY 3036 gatctcattgtagtcacataataaacctgcagagtgatatttccagagtgctcgataactgtta 3095
Db 2577 GATCTCATGTAGTCCATATATAACCTGCAGAGTGAATTTCCAGAGTGCCTCGAATACTTA 2636
QY 3096 attacatctcattagggctgaaaaagaatgacctagcttctgtatatacagctgtgtgct 3155
Db 2637 ATTACATCTCCATTAGGGCTGAAAGAATGACCTACGTTTCTGTATACAGCTGTGTGCT 2696

QY 3156 ttgatgtgtgttactgttacacagaagtgtgtgcactgaggtctgtgtgtccgta 3215
Db 2697 TTTGATGTTGTGTACTGTACACAGAGCTGTGTGCACGTGAGGCTCTGCGTGTGTCCGTA 2756
QY 3216 tggaaacctgtgtagccctgcgagttaagtaactgtgtccattcatattgtttacgttgaat 3275
Db 2757 TGGAAAGCCGTGTAGCCCTGCGAGTTAAGTACTGCTTCCATTGTTTACCGTGAAT 2816
QY 3276 ttctcccatggaatgttaagtaaaacttaagtgtttgtcatcataaatgtgaatact 3335
Db 2817 TTTTCTCCCATGGAATGTAAGTAAACTTAAAGTGTGTTCATCAATAATGTATACT 2876
QY - 3336 aaa 3338
Db 2877 AAA 2879

RESULT 7
AL133174/c
LOCUS
DEFINITION Human DNA sequence from clone RP3-470L14 on chromosome 20, complete sequence.
ACCESSION AL133174
VERSION AL133174.15 GI:8573761
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 113836)
AUTHORS Ramsay,H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT

On Jun 20, 2000 this sequence version replaced gi:8388434.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: , EMBL; Sw: , SWISSPROT; Tr: , TREMBL; Wp: , WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20
RP3-470L14 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: PCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RP3-470L14 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true right end of clone RP3-470L14 is at 113836 in this sequence. The true right end of clone RP1-155G6 is at 100 in this sequence.

FEATURES
Source

Location/Qualifiers
1. 113836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"

AUTHORS Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
TITLE Interaction of influenza virus NSI protein and the human homologue
of Staufen in vivo and in vitro
JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE 99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS Ortin,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,
Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN
FEATURES
source 1. 3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
36. 1526
/gene="staufen"
36. 1526
/gene="staufen"
/function="dsRNA binding protein"
/codon_start=1
/product="staufen protein"
/protein_id="CAB40082.1"
/db_xref="GI:4572588"
/translation="MKLGKKPMYKPPVDPYSRMOSTYNNMRGAYPRRYEYPPVPL
LYQVELSVGGQFNGKKTROAKHDAKALRLIÖNEPLPERLEVNGRESEENLNK
SEISOVEIALKRLPNFNEVARESGPRHMKNEVTKVSGEFVGESEKSKISKNA
AIAVLELKKLPPLPAVERVKPRIKKTKPIVKPQTSPEYGGINPISRLAÖIQAKK
EKEPEYTLITERGLPRREFVMQVKVNHTAEGTGNKKVAKRNAENMLEILGFKVP
ORSHQTRTÖVRGEDTHKETRGWKSNFLNLALGMENGTSNKEDFRPRLSHÖQLPA
GILPMVEVAÖAVGSÖGHHTKDFTRAAPNPAKATVTAMIARELLYÖSTPÄETILK
NNISSGHVPHGLTRPSEÖLDYLSRVÖGFÖVEYKDFPKNNKNEFVSLNCSÖPPLIS
HGIGKDVESHDMALNLIKLSLDÖQSTEMPRTGNGPMVCGRC"

BASE COUNT 891 a 693 c 699 g 783 t
ORIGIN

Query Match 83.6%; Score 2799.4; DB 68; Length 3066;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2870; Conservative 0; Mismatches 6; Indels 7; Gaps 6;

QY 459 gcataaccctactgtagaactaatgcaactgtgcatgaaacttggaaaaaaccaatgt 518
Db 1 GCATAACCCCTACTGTAGAACTAAATGCACTGTGCATGAAACTTGGAAAAAACCAATGT 60
QY 519 ataagcctgttaacccttactctcgcgtgcagtccacctataactacaacatgagagag 578
Db 61 ATAAGCCTGTTGACCTTACTCTCGGATGCACTCCACCTATACTACAACATGAGAGAG 120
QY 579 gtgcttatcccccgaggtaactttaccatttccagttccaccttacttatacaagtgt 638
Db 121 GTGCTTATCCCGAGGTACTTTTACCATTTCAGTTCACCTTACTTATCAAGTGT 180
QY 639 aacttctgtggagagacagcaatttaatgcaagaagaagaacagagctgcgaac 698
Db 181 AACTTTCTGTGGAGACACGCAATTTAATGGCAAAGAAAGACAGACAGGCTGCGAAC 240
QY 699 acgatgtctgtgccaagcgttgagatcctgcagaaatgagccctgcagagagctgtg 758
Db 241 ACGATGCTGTGCCAAGCGTTGAGGATCTGCAGAAATGAGACCCCTGCCAGAGAGCTGG 300
QY 759 aggtgaatggaagaatccgaagaagaanaatctcaataaatctgaaataagtcaagtct 818
Db 301 AGGTGAATGGAAGAAATCCGAAGAGAAATCTCAATAATCTGAATAATAGTCAAGTGT 360
QY 819 ttgagattgcacttaaacggaacttgcctgtgaaatttcgaggtgccccggagagatgtgc 878
Db 361 TTGAGATTGCACCTTAACCGGAACCTTGCTGTGAATTTTCGAGGTGGCCCCGGAGAGTGCC 420
QY 879 caccaccacatgaagaacttltgacaagaagttcgttgggagattgttggggaagtgt 938
Db 421 CACCCACATGAAGAACTTTGTGACCAAGCTTTCGTTGGGGAGATTGTGGGGGAAGGTG 480
QY 939 aagggaagaagaagatttcaagaanaaatgcgcgcatagctgttcttgagagctga 998

Db 481 AAGGAAAAGCAGAGAAGATTTCAAAGAAAAATGCCGCCATAGCTGTCTTGAGAGCTGA 540
QY 999 agaagttaccgccccctgcctgcagttgaaacyagtgtaaagcctagaatcmetaaagaacaa 1058
Db 541 AGAAGTTACCGCCCCCTGCTCGTAGTTGAACGAGTAAAGCCTTAGAATCAAAAAAGAAACAA 600
QY 1059 aaccatagtcgaagccacagacaagccagaatatgtgccagggatcaatccgatagcc 1118
Db 601 AACCATAGTCAAGCCACAGACAAGCCAGAATATGCGCAGGGGATCAATCCGATTAGCC 660
QY 1119 gactgcccagatccagcagcgaataaaggagaagagccagagttacacgtctctcacag 1178
Db 661 GACTGCCCCAGATCCAGCAGGCCAAAAAGAGAAGAGCCAGAGTACACGCTCTCACAG 720
QY 1179 agcgagcctcccgccgcgcgcaaggagtttgtatgcaggtgaaagttgaaaccacactg 1238
Db 721 AGCGAGGCTCCCGCCCGCCGAGGAGTTTGTGATGCAGGTGAAGTTGGAACCCACTG 780
QY 1239 cagaaggaacggcaccacaagaaggtgccaagcgcgaatgcaagccgagaacatgctgtg 1298
Db 781 CAGAAGGAACGGCACCAACAGAAGGTGGCCAAGCGCAATGCAGCCGAGAACATGCTGG 840
QY 1299 agatcctgtgttcaaaagtlcccgacggcgcaagcccaaccacccgactcaagtcagag 1358
Db 841 AGATCCTTGTTCAAAAGTCCCGCAGCG - -CAGCCACCAACCCCGCACTCAAGTCAAGAG 898
QY 1359 agaagacaccataaagaacacagggtgtgaaagaagaiaacc-tttttgaacctggc 1417
Db 899 AGAAGACACCCCAATAAGAAACCAAGGGATGGAAGAAAGTAACCTTTTGTGAACCGGC 958
QY 1418 tctgggat-gaaaatggtactagtaataaagagatgagttcagatgtccattataag 1476
Db 959 TCTGGGATGGAATAATGGGACTAGTAATAAGAGATGAGTTCAAGATGCCCTATCTAAG 1018
QY 1477 tcatacagcagctgcctgtctgtgaaattcttcccatgtgtgcccgaggtcgcacagcttagg 1536
Db 1019 TCATCAGCAGCTGCTGCTGGAATTTCTTCCCATGTGTGCCCGAGGTCGCCAGGCTGTAGG 1078
QY 1537 agttagtcaagagacatcacaccaagaatttaccagggcagctccgaatcctgtccaagc 1596
Db 1079 AGTTAGTCAAGGACATCACACCAAAAGATTTCACAGGGCAGCTCCGAATCCTGCCAAGCC 1138
QY 1597 caaggttaactgcattatagcccgagagttgtgtatgtggggcaactgcgccacagccga 1656
Db 1139 CACGCTAACTGCCATGATAGCCCGAGAGTTGTGTATGGGGCACCTCGCCACAGCCGA 1198
QY 1657 gaccattttaagaataaacatctcttcaaggccaagtlaccccatggaaccttcaagagacc 1716
Db 1199 GACCATTTTAAGATAAATACATCTCTTCAGGCCACGTACCACCATGGAACCTTCACGAGACC 1258
QY 1717 ctctgagcaactggaactatcttccagaagtlccaggaattccaggttgaatacaaaagactt 1776
Db 1259 CTCTGAGCAACTGCACTATCTTTCACAGAGTCCAGGATTCAGGTTGAATACAAAGACTT 1318
QY 1777 ccccaaaaaacaagaagaacgaatttgtatctcttatacaatgtctcctctcaagccactct 1836
Db 1319 CCCCCAAAACAACAAGAAGCAATTTGTATCTCTTATCAATTGCTCTCTCAGCCACTCT 1378
QY 1837 gatcagccatgtatcgcgaagatgtgagtlcctgcacatgatatgtcgtcgtgaacat 1896
Db 1379 GATCAGCCATGCTATCGGCAAGGATGTGAGTCTCGCCATGATATGCGCTGCAACAT 1438
QY 1897 cttaaagtgtgtctgagtttgaccaacaagtlacagagatgccaagaacaggaacag 1956
Db 1439 CTTAAAGTTGCTGTGAGTTGGACCAACAAGTACAGAGATGCCAAGAAGAGAAACCG 1498
QY 1957 accaatgtctgtgtgtggaggtgtcgaaccttctcgtgccatgaacatataaaatcc 2016
Db 1499 ACCAATGTCTGTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAACCATTTATAAATCC 1558
QY 2017 caacatatatactgaaaaataactgaaactgtttgaaatttggaaatttctgatactcca 2076
|||||

QY 1381 aggggatggaagaaagtaacctttttgaaacctggtctgggagtgaanaatgggactag 1440
 |||||||
 Db 1175 AGGGGATGGAAGAAAGTAACCTTTTGTGAACCTGGCTCTGGGATGAAGAAATGGGACTAG 1234
 QY 1441 taataaagagagatgagtccaagatgccttatctaagtcatcagcagctgctgtggaat 1500
 |||||||
 Db 1235 TAATAAAGAGAGATGAGTTCAGGATGCCCTTATCTAAGTCATCAGCAGCTGCCCTGTGAAT 1294
 QY 1501 tcttcccatggtgccgaggtcgcccaaggctgtagaagttagtcaaggacatcacacca 1560
 |||||||
 Db 1295 TCTTCCCATGCTGCCCGAGGTGCCCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAA 1354
 QY 1561 agatttaccagggcagctccgaaatcctgccaaggccaagtaactgccaatagaccg 1620
 |||||||
 Db 1355 AGATTTTACGAGGCGAGCTCCGAATCTGCGCAAGGCCACGGTAACGTGCATGATAGCCCG 1414
 QY 1621 agagtgtgtatgaggcacctcgccacagccgagaccattttaagaataacatctc 1680
 |||||||
 Db 1415 AGAGTTGTGTATGGGGGCGACCTCGCCACAGCCGAGACCATTTTAAAGAAATAACATCTC 1474
 QY 1681 ttcaggccaagtaacccatgacctctcagagagacctctgagcaactgagactatcttc 1740
 |||||||
 Db 1475 TTCAGGCCACGTACCCCATGAGACCTCTCAGAGACCCCTTGAGCAACTGACTATCTTTC 1534
 QY 1741 cagagtcacagagattccagttgaatacaaaagacttccccaaaaaacaacaagaacgaatt 1800
 |||||||
 Db 1535 CAGAGTCCAGGAGATTCAGAGTTGAATACAAAGACTTCCCCAAAAACAACAAGCAAGATT 1594
 QY 1801 tgtatctcttatcaatgtctcctctcagccacctctgatacagccatggtatcggaagga 1860
 |||||||
 Db 1595 TGTATCTCTTATCAATTGCTCTCTCAGCCACCTCTGTGATCAGCCATGGTATCGGCAAGGA 1654
 QY 1861 tgtgagtgccctgcatgatatgctgctgtaacatctaaagtgtgtctgagtga 1920
 |||||||
 Db 1655 TGTGAGTCTCCATGATATGGCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGA 1714
 QY 1921 ccaacaaagtaacagagatgccaagaacaggaacggaaccaatgtctgtgtggaggtg 1980
 |||||||
 Db 1715 CCAACAAAGTACAGAGATGCCAAGAACAAGAAACGACCAATGTCTGTGTGGAGAGTG 1774
 QY 1981 ctgaacctttctgccaatgaaccattataaaatcccaacatatatactgaaataactga 2040
 |||||||
 Db 1775 CTGAACCTTTTCTGGCCATGACCATTTATAAATCCCAACATATATACTGAAAAATACTGA 1834
 QY 2041 aactgcttttgaaaattctgaaattctgataacctcagtgggcgagagacagctggtga 2100
 |||||||
 Db 1835 AACTGCTTTGAAAAATTTGGAAATTTCTGATACCTCCAGTGGGCCGAGAGACACGGGTGTA 1894
 QY 2101 aaggaatgtggcagcagcagggaagaacaacagaacaacaagagcggtgtgcccgt 2160
 |||||||
 Db 1895 AAGGATGTGGCAGCAGCAGGGAAGACAAACAGAAACACAAGAGGCGGCTGTGGCCGGCT 1954
 QY 2161 ggaactgtctggtgtgtgtatgagccactcggtgacctggtgacctgacctgcaatag 2220
 |||||||
 Db 1955 GGACTGTGCTGGGTTGTGTGATGGCCACTCGGTGACCTGGCGGTCTCTACGCAATAG 2014
 QY 2221 cagctgctgtggggaagaagggtgccccagcagctgttctcccggaacaccagcaga 2280
 |||||||
 Db 2015 CAGCTGCCCTGTGGGAAGAGGGCTGCCACGCGCTGTCTCCCGGACACACGACAGA 2074
 QY 2281 tccacacccctgagcaccctcgltgtgtcttttttccctgtgtgaaagaagaacg 2340
 |||||||
 Db 2075 TTCCACACCCCTGGGACACCTCCGTGTTTGTGCTTTTCCCTGTGTGAAGAAGAAACG 2134
 QY 2341 gcaagaccccttctcaagctggtcactcagacacattggaacaacacctgagacgcat 2400
 |||||||
 Db 2135 GCACGACCCCTTCTCAAGCTGGCTCAGTACAGACATTGGGACAAACCTTGACAGCCAT 2194
 QY 2401 gccagagagagcctttagaccgccccagagctaaaaagcaccagagaaatcaaatgctt 2460
 |||||||
 Db 2195 GCCAGAGAGAGGCCCTTGTGACCGGGCCCGCAGAGCTAAAGACACGAGAAATCAATGCTT 2254

QY 2461 cctactcagcgtgaccccaacttttctagtggtgccacgccccaccacccctctgcagtlacc 2520
 |||||||
 Db 2255 CCTACTCAGCGGTGACCCCAACTTTTCTAGTGTGCGACGGCCCCACACACCTCTGCACTACC 2314
 QY 2521 cacacacacccactgcttctcctccaacagtgatctglatcttagttcatattt 2580
 |||||||
 Db 2315 CACACATCACCACTGCTTTCCTTCCCAACAGTGATCTGATTTCTTAGTTCAATTTT 2374
 QY 2581 ctttgattgatatgacacatatataaaatttcatctgagaaattctcaattgtatctag 2640
 |||||||
 Db 2375 CTTTGTATTGATATGACACATATATAAATTTTCATTGAGAAATTTCTCAATTGTATCTAG 2434
 QY 2641 ttaaatagcacagtttggnaactgtctgagactgacttataataatcacaaccgacaa 2700
 |||||||
 Db 2435 TTAATAAGCACAGTTTGGAAACTGTCTGAGACTGACTTTATCAATTAATTAACCGACAA 2494
 QY 2701 agatcatatccatgtgtatgtgttagacatttctatttcatctgactaaccagagagt 2760
 |||||||
 Db 2495 AGATCATATCCATGTGTATGTGTTAGACATTTTATTTTATTGACTTAACCCAGACAGT 2554
 QY 2761 ttcagtgatgcaaatltgtgtgccccctgtgttcagctgaaacagtlcctgactttcaaaa 2820
 |||||||
 Db 2555 TTCAGTGATGCAAAATGTGTGCCCTCTGTGTGAGTGAACAGCTCCGTGCACTTCAAAA 2614
 QY 2821 ccttgaataagtlctcccaagtlgtlataaatltgacaactttagaaattttaaactttaga 2880
 |||||||
 Db 2615 CCTTGAATAAGTCTCCACAGTGTATATAATTGACAAATTAGGAATTTTAAACTTTAGA 2674
 QY 2881 tgaatatttggttccatttttatttcatatttatttctgttaaagcaagagacttaa 2940
 |||||||
 Db 2675 TGATCATTTGGTTCCATTTTATTTATTTATTTTATTTGTAATGCAACAGACTTAA 2734
 QY 2941 tgaacttgatctctgtttttaaagattatataaaacacattgtgtatctatcatatgct 3000
 |||||||
 Db 2735 TGAACCTTGTATCTCTGTTTTAAAGATTATTAATAAACATTGTGTATCTATATATGGCT 2794
 QY 3001 ctgagagacttagcttctcactacactacagagatatcctccatgtagtccatataaacc 3060
 |||||||
 Db 2795 CTTGAGGACTTAGCTTTCCTACTACACTACAGATATGATCTCCATGTAGTCAATATAACC 2854
 QY 3061 tgcagagtgatlttccagagtgctcgatactgttaattacatctccatagggctgaaaa 3120
 |||||||
 Db 2855 TGCAGAGTGATTTTCCAGAGTGCTCGATACTGTTAATTACATCTCCATTAAGGGCTGAAAA 2914
 QY 3121 gaatgacctacgttctctgatatagcgtgtgtgtctttagtgtgtactgttacacag 3180
 |||||||
 Db 2915 GAATGACCTACGTTCTGTATACAGCTGTGTGCTTTGATGTGTGTACTGTACACAG 2974
 QY 3181 aagtggtgcaactgaggtctctgctgtgtgtccgtatggaacacctgtgagccctgaggt 3240
 |||||||
 Db 2975 AAGTGTGTGCACTGAGGCTCTGCGTGTGCTCCGTATGGAACCTGTGACCCCTCGAGT 3034
 QY 3241 taagtactgcttccatctcaattgtttagcgtggaatttctccccaatgaaatgtaagtaa 3300
 |||||||
 Db 3035 TAAGTACTGCTTCCATTCATTTGTTTACGCTGGAATTTTCTCCCATGGAATGTAAGTAA 3094
 QY 3301 aacttaagtgltgtcatcaataaagtgtlaataactaaaaa 3348
 |||||||
 Db 3095 AACTTAAGTGTGTTCATCAATAAATGTAATACTTAAAAA 3142

RESULT 6
 HSA132258
 LOCUS HSA132258 3066 bp mRNA PRI 14-MAY-1999
 DEFINITION Homo sapiens mRNA for stauflen protein, partial.
 ACCESSION AJ132258
 VERSION AJ132258.1 GI:4572587
 KEYWORDS stauflen gene; stauflen protein.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 3066)

[illegible]

D	b	241	ACGAGCC-----	247
Q	y	301	gigatctcgctcactgcaacctccacctccaggtcagcgatttcccacttcagctc	360
D	b	248	-----	247
Q	y	361	ccgataagctgagattacagagttatttaaccacttaacctctcagaactgacaagac	420
D	b	248	-----	247
Q	y	421	aacattgttcctggaacgccccttlttaaaaaagaagcataaccctactgtagaact	480
D	b	248	-----AGAAAGCATAAACCCTACTGTAGACT	274
Q	y	481	aaatgcactgtgcatgtgaaacttggaaaaaaacccaatgttaagccctgttgaccctc	540
D	b	275	AAATGCACCTGTGCATGAAACTTGGAAAAAACCAATGTATAAGCCTGTGACCCTTACTC	334
Q	y	541	tcgagatgcagtlccacctataactacaacatlgagagagtgctatccccgagtaact	600
D	b	335	TCGGATGCAGTCCACCCTATACTACAACATGAGAGAGAGTGTATCCCCGAGGTACTT	394
Q	y	601	ttaaccattccagttccaccttacttatacaagtgyaacttctgtggyagacaca	660
D	b	395	TTACCCATTTCAGTTCACACCTTACTTTATCAAGTGAACCTTCTGTGGAGAGACAGCA	454
Q	y	661	atttaatgycaaaagaaagacaagacagctgcgaaacacgatgtctgtgccaaagcgtt	720
D	b	455	ATTTAATGGCAAGGAAGACAGACAGACAGGCTGCCAAGACGATGCTGCTGCCAAGCGTT	514
Q	y	721	gaagatcctgcagaaatgagccctgcagagagcgtggaagtgaatggaagagaatccga	780
D	b	515	GAGGATCCCTGCAGAAATGAGCCCTGCCAGAGAGGCTGAGGTGAATGGAAGAGAAATCCGA	574
Q	y	781	agaagaaaaatctcaataaatctgaaataagltcaagltgttgagatgtgcaacttaacgaa	840
D	b	575	AGAGAAAAATCTCAATAAATCTGAAATAAGTCAAGTGTGTGAGATTGCACCTTAACGGA	634
Q	y	841	cttcgctgtgaatttcgaggtgccccggyagagtgtycccaaccacatgaagaacttct	900
D	b	635	CTTGCCCTGTGAATTTTCGAGGTGGCCCCGGAGAGTGGCCCAACCCACATGAAGAACTTGT	694
Q	y	901	gaaccaagtttcggttggygagttgttggyggaaggtgaagggyaaaaagcaagaagattc	960
D	b	695	GACCAAGGTTTCGGTTGGGAGTTTGTGGGGGAAGGTGAAGGGAACCAAGAGATTTC	754
Q	y	961	aaagaaaaatgccccatagctgttcttgagagctgaaagaagltaccgccccgtcctgc	1020
D	b	755	AAAGAAAAATGCCGCCATAGCTGTCTTGAGGAGCTGAAGAACTTACCGCCCTGCTGC	814
Q	y	1021	agttgaacgagtaaaagcctagaatcaaaaagaaaaaaccataagltcaagccacagac	1080
D	b	815	AGTTGAACGAGTAAAGCCTAGATCAAAAAAGAAAAAACCAACCCATAGTCAAGCCACAGAC	874
Q	y	1081	aagccccagaatatgcccagggtatcaatccgatltagccgaactgccccagatccagcagc	1140
D	b	875	AAGCCAGAAATATATGCCCCAGGGGATCAATCCGATTAGCCGACTGGCCCCAGATCCAGCAGGC	934
Q	y	1141	aaaaaagagaaagagccagagttacacgctctctcacagagcgaagcctcccgcgccgag	1200
D	b	935	AAAAAAGAGAAAGAGCCAGAGTACACGCTCCTCACAGAGCAGGCGCTCCCGCGCCGAG	994
Q	y	1201	ggaagtttgtatgcaaggtgaaggttggaaaaccacactgcagaaggaacggygcaaccaaa	1260
D	b	995	GGAGTTTGTATGCAGGTGAAGGTTGGAACACACACTGCAGAGAAGGAACGGGCACCAACAA	1054
Q	y	1261	gaaggtgyccaaagcgcaatbcagccgagaacatgtctggaatccttgyttcaaaagtccc	1320
D	b	1055	GAAAGGTGGCCAAAGCGCAATGACGCCGAGAACATGCTGGAGATCCTTGGTTCAAGTCCC	1114
Q	y	1321	gcaagcgcagcccaaccaaccgcacatcaagltcagagaggaagacaccataaagaacc	1380
D	b	1115	GCAGCGGCAGCCCAACCAACCCGCACTCAAGTCAGAGGGAAGACACCCATAAAGAAACC	1174

QY 1231 ccacactgcagaaggaacgqgcaaccaagaagtggccaagcgcaatgcagccgagaa 1290
|||||
Db 1068 CCACACTGCAGAAGGACGGGCACCAACAAGAAGGTGGCCAAAGCGCAATGCAGCCGAGAA 1127
QY 1291 catgctggaatccttggtttcaaaagtcgccgacgqgcagcccaaccaaccgcaactcaa 1350
|||||
Db 1128 CATGCTGGAGATCCTTGTTTCAAGTCCCGCAGCGCCAGCCCAACCCCGCACTCAA 1187
QY 1351 gtcagaaggagaagacacccaataaagaaccaggagtgaagaaaagtaaccttttga 1410
|||||
Db 1188 GTCAGAGGAGAAGACACCCATAAAGAAACCA - GGGATGGAAGAAAAGTAACCTTTTGA 1246
QY 1411 accctgcttggggatgaaaaatggactagttaataaaggatgagttcagatgcctta 1470
|||||
Db 1247 ACCTGGCTCTGGGATGAATAATGGGACTAGTAATAAAGAGATGAGTTCAGATGCCCTTA 1306
QY 1471 tctaagtcatacagcagctgcctgtctggaattcttcccatggtgcgccgagtcgccaggc 1530
|||||
Db 1307 TCTAAGTCATCAGCAGCTGCGCTGCGAATTCCTCCCATGCGTGCCCGAGGTCGCCCAGGC 1366
QY 1531 tgtagaagttagtcagaagacatacacaaagaatttaccaggcagctccgaatcctgc 1590
|||||
Db 1367 TGTAGGAGTTAGTCAAGGACATCACACCAAAGATTTTACCAGGGCAGCTCCGAATCCTGTC 1426
QY 1591 caagcgcaacgtaactgccaatgataagcccgagatgtgtlataggggcaacctcgccac 1650
|||||
Db 1427 CAAGGCCACGGTAACCTGCATGTAGATAGCCCGAGAGTTGTGTATGGGGCACCTCGCCAC 1486
QY 1651 agcgaagacatlttaagaataacatctcttcagcgcaacgtlaccatgacacctcac 1710
|||||
Db 1487 AGCCGAGACCATTTTAAAGATAAACACCTCTTCAGGCCACGTAACCCCATGAGCCTCTCAC 1546
QY 1711 gagaccctctgagcaactgagactatcttccagagtcacaggatltccaggttgatacaa 1770
|||||
Db 1547 GAGACCCCTCTGAGCACTGGACTATCTTCCAGAGTCCAGGGATTCAGGTGGAATACAA 1606
QY 1771 agacttcccaaaaaacaagaagcaaatgtglatctctatacaattgcctcctcagcc 1830
|||||
Db 1607 AGACCTCCCCAAAAACAACAAGCAAGCAATTTGTATCTTATCAATTGCTCTCTCAGCC 1666
QY 1831 acctctgatacagccatggtatcgcgaagatgtgagtcctgcataatgctgcgt 1890
|||||
Db 1667 ACCTCTGATCAGCCATGTATCGCAAGATGTGAGTCTCGCCATGATATGGCTGCGCT 1726
QY 1891 gaacatcttaaaagtgtctctgagttggaaccaaaagtlacagagatgccaagaacag 1950
|||||
Db 1727 GAACATCTTAAAGTTGCTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAG 1786
QY 1951 aaacggaccaatgtctgtgtgtgagagtgtgaaccttctctgccaatgaaccattata 2010
|||||
Db 1787 AAACGAGCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGCGCATGACCATTTATA 1846
QY 2011 aaatcccaacatatatactgaaaaatactgaaactgctltgaaaatttgaatttctgata 2070
|||||
Db 1847 AAATCCCAACATATATACTGAATAATACTGAAACTGCTTTGAAAAATTGGAATTTCTGATA 1906
QY 2071 cctccagttggccgagaagacacggtggtataaagatgtggcagcagcgagggaagacaac 2130
|||||
Db 1907 CCTCCAGTGGCCGAGAGACACGGTGGGTAAGGATGTGGCAGCAGCAGGGAAGACAAC 1966
QY 2131 agaaaacacaagagagcggtgtgccc - ggcgtgnaactgtctggggttgtgtgtagtggcc 2189
|||||
Db 1967 AGAAACACAAGAGAGCGGCTGTGGCCGGCTGACTGTGCTGGGGTTTGTGTGATGGCC 2026
QY 2190 acctcggtgacctgycggtccctaagcaatagcagctgcctgtggygaagaagggtgccc 2249
|||||
Db 2027 ACTCGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCCCTGTGGGAAGAAGGCTGCC 2086
QY 2250 agccagctggtctcctccggygaacacagcagatccaacccctggycaacctcgtgttgt 2309
|||||
Db 2087 AGCCAGCTGGTCTCCCGGACACCAAGCAGATCCACACCTGGGCACCTCCGTGTTGGT 2146
QY 2310 ctttttttccccctgtgtgaaaagaagaacgcaaccccccttctcaagctggtcactc 2369

Db 2147 CTTTHTTTTCCCTGTGTGAAGAAGAAACGGCACACCCCTTCTCAAGCTGCGCTCACTC 2206
QY 2370 agacacattgggacaaacctgtgacagccaatgycagagagaggcctltgacggccccaag 2429
|||||
Db 2207 AGACACATTTGGACAAACCCCTGGACAGCCATGCGCAGAGAGAGGCCCTTGACCGGCCAC 2266
QY 2430 agctaaaaagcaccagagaaaatcaaatgcttccctactccagcgtgacccaacttctagt 2489
|||||
Db 2267 AGCTAAAGCACCACGAGAGAAATCAATGCTTCCCTACTCAGCGTGACCCCACTTTTCTAGT 2326
QY 2490 gtgcacagggcccaaccactcctgcagtaaccaacaatcaaccaactgcttctctccaa 2549
|||||
Db 2327 GTGCCAGGGCCCCACACCTCCTGACGTACCCACACCATCACCACTGCTTCTTCCAA 2386
QY 2550 cagtgatctgtatcttagttcattatlttctttagatltgatatgacactatataaaat 2609
|||||
Db 2387 CAGTGATCTGTATCTTAGTTTCAATTATTTCTTTGATGATATGACACTATATAAAAT 2446
QY 2610 ttcaattgagaatttctcaattgtatctagttaaatagcacagtttgyaaactgtctg 2669
|||||
Db 2447 TTTCAATTTGAGAAATTTCTCAATTTGTATCTAGTTAAATAGCACAGTTTGAAACTTGCTG 2506
QY 2670 agactgacttlatcaataatctaaaccgacaagaatcatatccatggtgatatgttagac 2729
|||||
Db 2507 AGACTGACTTTATCAATATCTTAACCGACAAGATCATATCCATGTGTATGTGTTAGAC 2566
QY 2730 attttattcatltagactaacccaaggacagttlccagtatgcaaatgtgtgcctctg 2789
|||||
Db 2567 ATTTTATTTTCAATTGACTTAACCCAGGACAGTTTCAAGTATGCAAAATGTGTGCCCTCTGG 2626
QY 2790 tlccagctgaaacagtlccctggaacttcaaaaacctgaaataagtlcctccacagttgataa 2849
|||||
Db 2627 TTCAGCTGAACAGTCCCTGGACTTTCAAAAACCTTGAAATAAGTCTCCACAGTTGTATAA 2686
QY 2850 atlgacaattlagaaattttaaactltagatgataatltggtltccatttatttcaatt 2909
|||||
Db 2687 ATTGACAAATTTAGGAATTTTAAACCTTAGATGATCATTTGGTCCATTTTATTTCATT 2746
QY 2910 tttaatttltgttaatgcaaacaggactttaatgnaactltgatctgtltttaaagattat 2969
|||||
Db 2747 TTTACTTTTGTATTAATGCAACAGGACTTAATATGAACCTTGATCTGTGTTAAAGATTAT 2806
QY 2970 taaaaaacatltgtgtatctatatacatatggtcctctgaggacttagctltcaactacata 3029
|||||
Db 2807 TAAAAAACATTTGTGTATCTATACATATATGCGCTGTGAGGACTTAGCTTTCTACTACACTACA 2866
QY 3030 ggatatgatactccatgttagtcatalataaaacctgcagagtgatlttccagagtgctcgata 3089
|||||
Db 2867 GGATATGATCTCCATGTAGTCCATATATAAACCTGCAGAGTGATTTTCCAGAGTGTGATA 2926
QY 3090 ctgttaattacatctccatlagggctgaaagaatgacctaaglttctgtatacagctgt 3149
|||||
Db 2927 CTGTTAATTACATCTCCATTAAGGGCTGAAGAAAGTAAGACTTCTGTATACAGCTGT 2986
QY 3150 gtgcattttagatgtgtgttactgttacacagaaagtgtgcaactgagagctcgtggtg 3209
|||||
Db 3047 TCCGTAATGGAAGCCCTGTGTAGCCCTGCGAGTTAAGTACTGCTTCCATTCATTGTTACGC 3106
QY 3210 tccgtatgaaaaacctgtgtagccctgcgagttlaagtaactgtcttccattcaattgttaagc 3269
|||||
Db 3270 tggaaatttctccccaatgaaatgtaagtaaaacttaagtgltgtlcatcaataaatggt 3329
|||||
Db 3107 TGGAAATTTTCTCCCATGGAATGTAAGTAAGACTTAAGTGTGTTGTATCAATAATGCT 3166
QY 3330 aatactaaaaaaaanaa 3348
|||||
Db 3167 AATACTAAAAAANAANA 3185
RESULT 5

QY	3263	tttagcgtggaattttcttcccatggaatgtgaagtaaaacttaagtgttgcacat	3322
Db	3421	TTTACGCTGGAATTTTCTCCCATGGAATGTAAGTAAACTTTAAGTGTTCATCAAT	3480
QY	3323	aaatgtaataactaaaaa	3348
Db	3481	AAATGTAATACTAAAAA	3506
RESULT	4		
LOCUS	AK025519		
DEFINITION	AK025519 Homo sapiens CDNA: FLJ21866 fis, clone HEP02379, highly similar to AF061938 Homo sapiens staufen protein (STAU) mRNA.		
ACCESSION	AK025519		
VERSION	AK025519.1	GI:10438059	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens hepatoma cell_line:HepG2 CDNA to mRNA, clone_11b:HEP clone:HEP02379.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (sites)		
TITLE	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
JOURNAL	NEDO human cDNA sequencing project		
REFERENCE	Unpublished (2000)		
AUTHORS	2 (bases 1 to 3194)		
TITLE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
JOURNAL	Direct Submission		
COMMENT	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
FEATURES	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology; CDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
source	Location/Qualifiers		
	1..3194		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="HepG2"		
	/cell_type="hepatoma"		
	/clone="HEP02379"		
	/clone_11b="HEP"		
	/note="cloning vector pME18SFL3"		
	1..3194		
	/note="highly similar to AF061938 Homo sapiens staufen protein (STAU) mRNA"		
BASE COUNT	894 a 787 c 732 g 781 t		
ORIGIN			
Query Match	88.4%; Score 2961; DB 34; Length 3194;		
Best Local Similarity	95.4%; Pred. No. 0;		
Matches 3167; Conservative	0; Mismatches 15; Indels 137; Gaps 6;		
QY	33	cttcagcgttgcgc--ggcgcgtgcgcgtctctctcgcgtcccgcttccttgacgcc	90
Db	1	CTTCAGCGTTGCGCGCGCGCGGCTGCGGCTCTCTCTCGGCTCCCGCTTTGACCGCC	60
QY	91	tcccccccgccggcgcgcgccgcctctctcaacggccactccgctcttcctccct	150
Db	61	T-CCCCCGCGCGCGCGCGCGCTCTCTCCACGGCCACTCCGCTCTTCCCTCCCT	119

QY	151	tgcgcccttcttcctcctcccttttllctctctcctcccctctcgccgaccgccccca	210
Db	120	TGCTCCCTTCTTCCTCCTCTTTTTCCTTCTTCCCTCCCTCCCGGCCACGGCCCA	179
QY	211	ggacccgagccgagggaagagtcggagagcagccagccagccagtgtgatttg	270
Db	180	GGACC GCCCGGGGAGCAGACTCGAGCAGCAGC-----	215
QY	271	ctctgtgccagcgtgagtgacgtgcgtgatctcgcgtactgcactccaactc	330
Db	216	-----	215
QY	331	ccaagtcagcgaatttcccacttcagcctcccgaatagctgagatlacagaattattaa	390
Db	216	-----CAGAGTTATTTAA	228
QY	391	ccacttaacctctcagaactyaaacaagaacacatigtctctgyaaogccctcltta	450
Db	229	CCAATTAACTCTCAGAACTGAACAAGAACAACATGTGCTTGAAACGCCCTTTTTAA	288
QY	451	aaaagaaagcataaccctactgttagaactaatgcatgtgcatgaaactltyaaaaa	510
Db	289	AAAAGAAAGCATAAACCCCTACTGTAGAACTAATAATGCACGTGCATGAAACTTGAAAAA	348
QY	511	accaatgtataagcctgttgacccttactctcgcgatgcagtcacctaataactacaacat	570
Db	349	ACCAATGTATTAAGCCTGTGAACCTTACTCTCGGATGCAGTCCACCTATACTACACAT	408
QY	571	gagaggaagtgcctatcccccgaggtaactlltacccaatlccagtlccaccttacttta	630
Db	409	GAGAGGAGGTGCTTATCCCCGAGGTACTTTTACC CATTTCCAGTTCCACCTTTACTTTA	468
QY	631	tcaaagtgyaaccttctgttgygaagcacgcaattlaatygc aaagyaagacaagaacagc	690
Db	469	TCAA GTGAACTTTCGTG GAGAGGACAGCAATTTAATGCCAAAGGAAGACAAGACAGACGC	528
QY	691	tgcgaaacaacgatgtcgtctgc caaagcgtltgagatcctgcagaatlygccccctgcaga	750
Db	529	TGCGAAACACGATGCTGCTGCCAAAGCGTTGAGGATCTCGCAGAATGAGCCCCCTGCCAGA	588
QY	751	gagcgtgaggtgaatgtgaagagaatccgaagaagaanaatctcaataatctgaanaag	810
Db	589	GAGGCTGGAGGTGAATGGAAGAGAAATCCGAGAAGAAATCTCAATAATCTGAANAATAG	648
QY	811	tcaaagtgttgagattgcactttaaacgygaacttgccctgtgaaatlctgagtygccccgyga	870
Db	649	TCAA GTGTTGAGATTGC ACTTTAAACGGAACTTGCCCTGTGAATTTTCGAGGTGCGCCGGGA	708
QY	871	gagtggcccaaccacatatgaagaactltgtgaccaagtltcgyltgygagtltgtgyg	930
Db	709	GAGTGGCC CACCCACATGAGAACTTGTGACC CAAGTTCGGTTGGGAGTTGTGGG	768
QY	931	ggaagtgtaagggaaaaagcaagaagatttcaagaagaaatgcccgaatagctgttctga	990
Db	769	GGAAGGTGAAGGGA AAAAGCAAGAAGATTTCAAAGAAAAATGCCGCATAGCTGTTCTTGA	828
QY	991	ggagctgaagaagttaacggccccctgcctgcagtltgaacgagtaaagcctagaatcaaaaa	1050
Db	829	GGAGCTGAAGAAGTTACCGCCCTGCTCTGCAGTTGAACGAGTAAAGCCTAGAAATCAAAAA	888
QY	1051	gaaaaacaaaccatagtcacagccaacagacaagcccagaatatygccaagygatcaatcc	1110
Db	889	GAAAAACA AAA -CCATAGTCAAGCCACAGACAAGCCCAGATATATG GCGCAGGGGATCAATCC	947
QY	1111	gattagccgactggcccagatccagcagcagycaaaaaagggaagagccagagtacacgct	1170
Db	948	GATTTAGCCGACTGGCC CAGATCCACAGCAGCAGCAAAAAAAGAGAGAGCCAGAGTACACGCT	1007
QY	1171	cctcacagagcagagcctcccgcgcgcgcaagygagtttctgtatgcaggtgaaggttggaaa	1230
Db	1008	CCTCACAGAGCGAGG CCTCCCGCGCCGCGCAGGAGGATTTGTGATGCAGGTGAAGGTTGGAAA	1067

Db 1201 ATCAAAAAGAAAACAAACCCATTAGTCAAGCCACAGACAAAGCCAGAAATATGGCCAGGGG 1260
Qy 1103 atcaatccga ttagccga tggccagatccagcag gcaaaaaagagagagcagcag 1162
Db 1261 ATCAATCCGATTAGCCGACTGGCCAGATCCAGACGCAAAAAAGAGAGAGCCAGAG 1320
Qy 1163 tacacgctcctcacagagcgaggcctcccgcgccgcaaggagtgtgtatgcaggtgaag 1222
Db 1321 TACACGCTCCTCACAGAGGAGGCCCTCCCGCGCCGAGGGAGTTGTGATGCAGGTGAAG 1380
Qy 1223 gttggaaccacactgtcagaagaggaacg ggcacccaagaagtgtgccaaagcgcaatgca 1282
Db 1381 GTTGGAAACACACACTGCACAGAAGAACGGGCACCAAGAAGGTGGCCAAAGCCAATGCA 1440
Qy 1283 gccgagaacatgtctgagatccttgytltcaaag tcccgcagcggcagcccaaaccc 1342
Db 1441 GCCGAGACATGCTGAGATCCTTGTTCAAAGTCCCGCAGCGGCAGCCCAACCAACCC 1500
Qy 1343 gcactcaagtcagagagagaaagacaccataaagaaaccagggatgtgaagaaagtaacc 1402
Db 1501 GCACTCAAGTCAGAGAGAAAGACACCCATAAAGAAACCAGGGATGGAAGAAAGTAACC 1560
Qy 1403 tttttgaacctgctctgyggatgtaaaa tgygactagtaataaagagatgaqtlcag 1462
Db 1561 TTTTGTGAACCTGGCTGTGGGATGAAATGGGACTAGTAATAAGAGATGAGTTCAGG 1620
Qy 1463 atgccttatctaagtcacagcagctgcctgtggaatcttcccatgtgtcccgagtc 1522
Db 1621 ATGCTTATCTAAGTCATCAGCAGCTGCCCTGCTGGAATTCTTCCCATGTGTCGAGGTC 1680
Qy 1523 gcccaggtgtagagtagtcaagacatcacaccaagaatlttacagggcagctccg 1582
Db 1681 GCCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGATTTTACGAGGCGAGCTCCG 1740
Qy 1583 aatcctgccaagggccaggtlaactgcatgtacgtacgcccagagttgtgtatgygggcacc 1642
Db 1741 AATCTGCCAAGGCCACGGTAAC TGCCATGATGAGCCGAGAGTTGTGTATGGGGCACC 1800
Qy 1643 tcgcccacagccgagacca ttttaagaataacatctcttcagggccagttaccccatgga 1702
Db 1801 TCGCCACAGCCGAGACCATTTTAAAGATATACATCTCTTCAGGCCACGTACCCCATGGA 1860
Qy 1703 cctctcaagagaccctctgagcaactggaatcttccagagtcacaggaatccaggt 1762
Db 1861 CCTCTACGAGACCCCTGTGAGCAACTGACATATCTTCCAGAGTCCAGGGATTCAGGTT 1920
Qy 1763 gaatacaaaagacttccccaaaaaacaacaagaagaa tltgtatctctatacaattgtccc 1822
Db 1921 GAATACAAAGACTTCCCAAAAAACAAGAAGAAATTGTATCTCTATCAATTGCTCC 1980
Qy 1823 tctcagccacctgtgatcagccatgtgtatcggaagatgtgtgagttccgtgcatgtatg 1882
Db 1981 TCTCAGCCACCTCTGATCAGCCATGATGTATCGGCAAGGATGTGAGTCCCTGCCATGATATG 2040
Qy 1883 gctgcgtgaaacatcttaaaag tttgtctgtcgtatgtgacccaacaagaatcacagatgcca 1942
Db 2041 GCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGGACCAACAAGTACAGAGATGCCA 2100
Qy 1943 agaacaggaaaacggaacaatgtctgtgtgtggaaggtgtctgaaccttctctgcatgaa 2002
Db 2101 AGAACAGGAAACGGACCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAA 2160
Qy 2003 ccaatataaaatcccaacataatactgaaaaaactgtaactgtcttgaaaaatltgnaat 2062
Db 2161 CCATTATAAATCCCAACATATATACTGAAAAATACTGAAACTGCTTGAAAAATTTGGAAT 2220
Qy 2063 tctctatacctccagtggtggccgagagacacagtggtggtlaaagatgtgtggcagcaggy 2122
Db 2221 TTCTGATACCTCAAGTGGGCCGAGAGACACAGGTTGGGTTAAAGATGTGGCAGCAGAGGG 2280
Qy 2123 aagacaacagaaacacaaaggaggcggtgtgtgcccggctggaactgtgtcgtgggtttgtgt 2182
|||||

Db 2281 AAGACAACAGAAACACAGAGAGCGGCTGTGGCCGGCTGCACTGTGCTGGGTTTGTGT 2340
Qy 2183 gatggccaactcgtgtacctggtgcgttccctacgcgaatagcagctgcgtgtggygaagag 2242
Db 2341 GATGCCCACTCGGTGACCTGGCGGTTCCCTACGCAATAGCAGCTGCCCTGTGGGGAAGAAG 2400
Qy 2243 gctgcccaagcagctgttctcccggtgtgaagaagaacggaacggaaccccttccaagctg 2302
Db 2401 GCTGCCAGCCAGCTGTTCTCCCGGACACACAGATCCACACCCCTGGCACCTCCGT 2460
Qy 2303 gtttgctcttlttccctgtgtgaagaagaacggaacggaaccccttccaagctg 2362
Db 2461 GTTGTGCTTTTTCCTTCCCTGTGAAAGAAGAACGGCACGACCCCTTCTCAAGCTGG 2520
Qy 2363 ctcactcagacacatltgggacaacccttgacagccaatgccaagagagagcctltgaccg 2422
Db 2521 CTCACCTCAGACACATTTGGACAAACCCCTGGACAGCCATGCCAGAGAGAGGCTTGACCG 2580
Qy 2423 gccccagaagctaaaagcaccagaagaatacgaatgtcttccactactcagcgtgacccaact 2482
Db 2581 GCCCCAGAGCTAAAAAGCACACAGAGAAAAATCAAAATGCTTCTTACTCAGCGGTGACCACTT 2640
Qy 2483 ttctagltgtccacgcccccaacctctctgcagtaaccacacatcacactgtctct 2542
Db 2641 TTCTAGTGTGCCACGGCCCCACCACTCCTGCAGTACCCACACATCACACTGCTTCT 2700
Qy 2543 ctccaacagtgatctgtatcttagtltcatatlttctltgattgatgtacacat 2602
Db 2701 CTTCACAACAGTATCTGTATTTCTTAGTTCAATATTTCTTTGATGTATGACACTAT 2760
Qy 2603 ataaaaatltcatlttgagaatltctcaatltgtatctagttaaatagcacaglttgyaac 2662
Db 2761 ATAAATTTTCATTTGAGAAATTTCTCAATGTATCTAGTTAAATAGCACGTTTGGAAC 2820
Qy 2663 ttgtctgagactgactltcaataatactcaaccgaagaatcatatccaatgtgtatgt 2722
Db 2821 TTGCTGAGACTGACTTTATCATATATCTAACCCGACAAGATCATATCCATGTGTATGTG 2880
Qy 2723 gttagaacatlttlatltcatgtactaaaccaggaaglttcagltgatgcaaatgtgtgc 2782
Db 2881 GTTAGACATTTTATTTTCATGTAGCTAACCCAGACAGTTTCAGTGATGCAAAATGTGTGC 2940
Qy 2783 cctctggttcagctgaaacagctcctggaacttcaaaaaacctgtaataagcttcccacagt 2842
Db 2941 CCTGTGTTCAAGCTGAACAGTCTTGACTTTCAAAAAACCTTGAATAGCTTCCACAGT 3000
Qy 2843 tgtataaatltgacaatlttaggaatlttlaaactttagatgatcatltgttccatlttla 2902
Db 3001 TGTATAAATTTGGACAATTTAGGAATTTTAAACTTTAGATGATCATTTGCTCCATTTTAA 3060
Qy 2903 ttcaatlttlatlttgttcaatgtcaaacaggaacttlaaactltgaatctgttcttlaa 2962
Db 3061 TTTCAATTTTATTTTGTATATGCAAAACAGGACTTAAATGAACCTTGTGATCTGTTTAA 3120
Qy 2963 agatlatlaaaaaacatltgtlatcatalacatagctcctlgagactagctltcaacta 3022
Db 3121 AGATTAATTAATAAACATTTGTATCTATACATATGCTCTTGAGGACTTACCTTCACTA 3180
Qy 3023 cactacaggaatagatctcatgttagtccatalaataacctgcaagatgtatltccagagt 3082
Db 3181 CACTACAGGATATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGAATTTCCAGAGTG 3240
Qy 3083 ctcgatactgttlaatatacttccatlttaggctgtgaaaagaatgacctagttctgtata 3142
Db 3241 CTCGATACTGTTAATTTACATCTCCATTAGGGCTGAAGAAATGACCTTCTGTGATA 3300
Qy 3143 cagctgtgtcttcttgatgttgtgttactgtacacagaagtgtgtgcactgagctctg 3202
Db 3301 CAGCTGTGTTGCTTTTGTATGTGTGTACTGTACACAGAAGTGTGTGCACGAGGCTCTG 3360
Qy 3203 cgtgtgttccgtatgtgaaaacctgttagccctcgaglttaagtaactgtcttcaatctg 3262
Db 3361 CGTGTGTCGATGAAAAACCTGTGTAGCCCTGCGAGTTAAGTACTGCTTCCATTCATTG 3420
|||||

Db	3110	TAAGTACTGCTTCATTCATTGTTTACGCTGGAATTTTCTCCCATGGAATGTAAGTAA	3169
Qy	3301	aacttaagtgttgcatacaataaatggtataactataaaaaaaaaa	3348
Db	3170	AACTTAAGTGTTCATCATTAATAATGTAATACTTAAAAA	3217
RESULT	3		
LOCUS	AF061939		
DEFINITION	AF061939	3506 bp mRNA PRI 04-MAR-1999	
ACCESSION	AF061939	Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced, complete cds.	
VERSION	AF061939.1	GI:4335948	
KEYWORDS		human.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 3506)	
TITLE		Wickham, L., Duchaine, T., Luo, M., Nabi, I.R. and DesGroselliers, L. Mammalian stauflen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum	
JOURNAL		Mol. Cell. Biol. 19 (3), 2220-2230 (1999)	
MEDLINE		99147057	
REFERENCE		2 (bases 1 to 3506)	
AUTHORS		DesGroselliers, L., Wickham, L. and Luo, M.	
TITLE		Direct Submission	
JOURNAL		Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada	
FEATURES		Location/Qualifiers	
source		1. .3506	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="20"	
		/map="20q13.1"	
		1. .3506	
		/gene="STAU"	
		409. .2142	
		/gene="STAU"	
		/function="RNA binding protein"	
		/note="alternatively spliced transcript"	
		/codon_start=1	
		/product="stauflen protein"	
		/protein_id="AAD17531.1"	
		/db_xref="GI:4335949"	
		/translation="MSQVQVQVQNPASALSGSILNKNSLLSQPLMSTPSTSSLPSS	
		ENAGRPIONSAIPASITSTSAAESITPEVLNALCMKLGKPPMKPDPYSRMOST	
		YNNMRGAGAYPRFYFPVPVPLLYQVELSVGGQFNGKGTROAKHDAKALRIL	
		QNEPLPERLEVNGRESEENLNKSEISQVEIALKRLPNVFEVARESGPHMKNFTV	
		KVSVGEFVGEGEKSKKISKNAALIAVLEELKLLPLPAVERVKPRIKKTPIVQPQ	
		TSPEYGGQINPISRLAQIOQAKKEKEPEYTLTERGLPRREFVMQVGNHTAEGTG	
		TNKKVAKRNAENMLEILGFKVPQRPQTPRALKSEKTPIKKPGDGKRVTFEFGSGD	
		ENGTSNKEDFERMPYLSHOQLPAGILPMVEVAQAVGSQGHHTKDFTRAAPNPAKAT	
		VTAMIARELLYGTSPTAETILKNNISSGHVPHGPLTRPSEQLDLSRYQGFQVEYKD	
		FPKNNKNEEYSLINCSSQPLPLISHGIGKVESCHDMAALNLIKLLSELDOOSTEMPR	
		GNGPMSVCGRC"	
BASE COUNT	965 a	875 c	790 g 876 t
ORIGIN			

[illegible]

QY	121	ctccacgagccactccgcgcctctccctccctctgctccctctctctccctctccctcttctcct	180
Db	121	CTCCACGGCCACTCCGCCCTCTTCCCTCCCTCTGTCCTTCTTCTCTCTCCCTTTTTCCTT	180
QY	181	cttccttccctctctcgcgcgcacgcgcacagaccgcgcgcgggagacagtcgcgagc	240
Db	181	CTTCCTTCCCTCCCTCCGCCGCCACCGCCAGACCGCGCGGGGACGAGTCCGGAGC	240
QY	241	agcagccag-----	249
Db	241	AGCAGCCAGAGTTTATTAAACCACCTTAACCTCTCAGAACTGACAACAAGACAACATTTGTCC	300
QY	250	-----cagcagccagctgagcttctgctct-----	275
Db	301	TGGAACGCCCTCTTTTAAAAAAGGTAGAACTTTAGACTTTCATAGCACTGAATTAACCTG	360
QY	276	-----gtcgcgccagctgagtg	293
Db	361	CACCTGAAGCTGTTTTACCTGCATTTGTTCACCTTTGTGTTGAAGTGACCATGTCTCAAGTT	420
QY	294	cagtgcgctgatactcgcgtcactgcaacctccacctcccaagtcagcgatlttccactt	353
Db	421	CAAGTGCAAGTTGAGAACCCCATCTGCTCTCTCTCAGGGAGGCCAATTACTGAACAGAAC	480
QY	354	cagcctcccgataagctgagattacagagttatlaaccact-----	395
Db	481	CAGTCTCTTCTTCACAGCCTTTGATGAGTATTCCTTCTACTACTAGCTCTCTGCCCTCT	540
QY	396	-----taacctcagacactgnaacaaagacacacatgttccctgnaagccct	442
Db	541	GAAATGCAGGTAGACCCCATTCAAAACTCTGCTTACCCTCTGCATCTATTACATCCACC	600
QY	443	ctlttlaaanaagaaagcataaacccctactgttagaactaaatgcaactgtgcatgnaact	502
Db	601	AGTGCAGCTGCAGAAAGCATTAACCCCTACTGTAGAACTAAATGCACCTGTGCATGAACTT	660
QY	503	ggaaaaaaaccaatgataaagcctgttgaacccctactctcgatgagtcagtcacactataac	562
Db	661	GGAAAAAAACCAATGTATTAAGCCTGTGTACCCCTTACTCTCGATGCAGTCCACCTATTAAC	720
QY	563	tacaacatgagagaggtgtctatcccccgaggtacttaccacatltccagtlccacct	622
Db	721	TACAACATGAGAGAGGCTGCTTATCCCCGAGGTACTTTTACCATTTCAGAGTTCACCT	780
QY	623	ttactttatcaagtgyaacttctgtgagagagacagcaatttaatgccaagaagaca	682
Db	781	TTACTTTATCAAGTGAAGACTTCTGTGGGAGGACAGCAATTTATGGCAAAAGAAAGACA	840
QY	683	agacaggtctcgaaacacacgatgtctgtcgccaaagcgttgagatccctgcagaatgagccc	742
Db	841	AGACAGGCTGCCAAACACGATGCTGCTGCCAAAGCGTTGAGGATCCCTGCAGATGAGCCC	900
QY	743	ctgcacagagagctgygaggtgtaatggaagagaatccgaagaagaanaatcctaataatct	802
Db	901	CTGCCAGAGAGGCTTGAGAGGTGAATGGAAGAGAATCCGAAGAAGAAAATCTCAATAAATCT	960
QY	803	gaaataagtcaagtlttgagattgcaacttaaacggaacttgccctgtgaatttcgagatg	862
Db	961	GAAATAAGTCAAGTGTGTGAGATTGACATTAAACGGAACCTGCTGTGAATTTGAGGTG	1020
QY	863	gcccgggagagtgcccacccccacatgaaagaaactltgtgaccaaaggttgcgttgaggag	922
Db	1021	GCCCCGGAGAGTGGCCCCACCCACATGAAGAACTTTGTGACCACAAGCTTTCGGTTGGGGAG	1080
QY	923	tttgtggyggaagtgyaaggygaaagcaagaagatttcaagaaaaaatgcccgcataagct	982
Db	1081	TTTGTGGGGGAAGGTGAAGGGAAGAAAGCAAGAGATTTCAAAAGAAAATGCCGCATAGCT	1140
QY	983	gttccttgagagctgaaagaagttacgcgccccctgctgcagttgaaacygtaagcctaga	1042
Db	1141	GTTCTTGAGGAGCTGAAGAAGTTACCGCCCCCTGCCCTGCAGTTGAACGAGTAAAGCCTAGA	1200
QY	1043	atcaaaaaaagaaaaaaccataagtcagcaccagacagaacaagcccaagaatatatgcccaggg	1102

QY 1081 aagccccaatatgcccaggggatcaatccgattagccgactgcccagatccagcagc 1140
|||||
Db 950 AAGCCCAAGATATGCCCCAGGGGATCAATCCGATTAGCCGACTGGCCACAGTCCAGCAGGC 1009
QY 1141 aaaaaaggagaaggagccagagtagacgctcctcaacagagcgagcctccgcgcgcag 1200
|||||
Db 1010 AAAAAAGGAGAGAGCCAGAGTAGACGCTCCTCACAGAGCGAGGCGCTCCGCGCGCAG 1069
QY 1201 ggaagtttgtatgtagcaggtgaaggtltggaaccacactgcagaaggaagcggcaccaaca 1260
|||||
Db 1070 GGAGTTTGTATGCAGGTGAAGGTTGGA AACCACTGCAGAGAAGAACGGGCACCAACAA 1129
QY 1261 gaaagtggccaagcgcaatgtagccgagaaacatgctgagatcccttgttcaagtccc 1320
|||||
Db 1130 GAAGGTGGCCAAAGCCCAATGCAGCCGAGAACATGCTGAGATCCTTGGTTCAAGTCCC 1189
QY 1321 gcaagcgagccccaacaaacccgcactcaagtcaagagagaagacacccataaagaacc 1380
|||||
Db 1190 GCAGCGGCAGGCCCAACCAACCCGCACCTCAAGTCAAGAGAGAACACACCCATTAAGAAACC 1249
QY 1381 agggagatggaagaaagtaaccttttttgaacctgctcgtggatgaaaaatgggactag 1440
|||||
Db 1250 AGGGGATGGAAGAAAGTAACCTTTTGAACCTGGCTCTGGGATGAATAAGGACTAG 1309
QY 1441 taataaagagatgagttcaggaatgccttatctaagltacagcagctgcctgtgaat 1500
|||||
Db 1310 TAATAAAGAGGATGAGTTCAGGATGCCCTTATCTAAGTCATCAGCAGCTGCCTGTGAAT 1369
QY 1501 tcttcccatggtgcccgaaggtcgcccagcgctgtagaggttagtcaaggacatcacacca 1560
|||||
Db 1370 TCTTCCCATGTTGCCGAGGTGCCCCAGGCTGTAGAGTTAGTCAAGACATCACACCAA 1429
QY 1561 agattttaccagggcagctccgaatcctgcccagggccaagtgtaactgccaatgtagccg 1620
|||||
Db 1430 AGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGGTAACCTGCCATGATAGCCCG 1489
QY 1621 agaagttgtgtatggyggcacctcgcccacagccgagaaccaatttaagaaataacatctc 1680
|||||
Db 1490 AGAGTTGTGTATGGGGGCACCTCGCCACAGCCGAGAACCATTTTAAAGATAACATCTC 1549
QY 1681 ttcagggccaagtcacccaatggaacctcagcaggaacctctgagcaactgactatcttc 1740
|||||
Db 1550 TTCAGGCCACGTACCCCATGGAACCTCTCACGAGACCTCTGAGCAACTGACTATCTTTC 1609
QY 1741 cagaagtcagggattccaggttgaatataaagaacttccccaaaaacaacaagaagaatt 1800
|||||
Db 1610 CAGAGTCCAGGATTCAGGTTGAATACAAAGACTTCCCCAAAACAACAAGACGAATT 1669
QY 1801 tgtatccttatacaatgtctcctcctcagccacctctgtatcagccatgttatcggaaga 1860
|||||
Db 1670 TGTATCTCTTATCAATGCTCCTCTCAGCCACCTCTGATCAGCCATGTTATCGGCAAGGA 1729
QY 1861 tgtgagtcctgccaatataatgctgcgtggaacatcttaagttgctgtgagttga 1920
|||||
Db 1730 TGTGGAGTCTCTGCCATGATATGGCTGGCTGGAACATCTTAAAGTTGCTGTGAGTTGA 1789
QY 1921 ccaacaagttacagagatgccaagaagaacggaacgaacaaagtctgtgtggyaggtg 1980
|||||
Db 1790 CCAACAAAGTACAGAGATGCCAAGAAGAACGGAACGCAATGTCTGTGTGGAGGTG 1849
QY 1981 ctgaaccttttctgcatgaaccaatataaatacccaacatatataactgaaataactga 2040
|||||
Db 1850 CTGAACCTTCTCTGGCCATGAACCATTTATAAATCCCAACATATATACTGAATAACTGA 1909
QY 2041 aactgctttgaaaatttgyaaatttctgatacctcagtggygcgagagaacacggtggt 2100
|||||
Db 1910 AACTGCTTTGAAAATTTGGAATTTCTGATACTCTCAAGTGGGCCGAGAGACACGGTGGTA 1969
QY 2101 aaggaatgtgagcagcaggaaggaacaaacagaacaaacaagagagcggtgtgcygct 2160
|||||
Db 1970 AAGGATGTGGCAGACGACGGAAGACAAACAGAAACACAAGAGAGCGGCTGTGCGCGCT 2029
QY 2161 ggaactgtgctggygttctgtgtgtatgtagccactcggtgacctgcygctccctacgcaatag 2220

Db 2030 GGACTGTGCTGGGGTTGTGTGATGGCCACTCGGTGACCTGGCGGTCCCTACGAATAG 2089
QY 2221 cagctgctgtggygaagagggctgcccagccagctggttctccggygacaaccagcaga 2280
|||||
Db 2090 CAGCTGCTGTGGGGAAGAGGGCTGCCCCAGCCAGCTGTCTCCCGGAGACCAACGAGA 2149
QY 2281 tccaacacctggygcacctccggtgttgtcttttttccctgtgtgaagaagaacg 2340
|||||
Db 2150 TCCACACCTTGGGCACCTCCGTGTGCTCTTTTTCCTGTTCCCTGTGTGAAGAGAAGAACG 2209
QY 2341 gcaagaccccccttcaagctggtcactcagacacatltggyacaacacccctgagcagcat 2400
|||||
Db 2210 GCACGACCCCTTCTCAAGCTGGCTCACTCAGACACATTTGGACAAACCTTGACAGCCAT 2269
QY 2401 gccagagagagcgctltgaccggyccccagagctaaagcaccagayaaatcaaatgtcct 2460
|||||
Db 2270 GCCAGAGAGAGGCTTTGACCGGCCCCAGAGCTAAAGCACAGAGAAATCAAAATGCTT 2329
QY 2461 cctactcagcgtgacccaacttcttagtgtgccaaggycccccaccacctcctgcagtacc 2520
|||||
Db 2330 CCTACTCAGCGTGACCCCACTTTCTAGTGTGCCACGGCCCCACCACTCCTCGAGTACC 2389
QY 2521 cacaccatcaccaactgcttctcttccaacagtgatctgtatcttagtltcaatltt 2580
|||||
Db 2390 CACACCATCACCACTGCTTCTCTTCCAACAGTGATCTGATTTCTTAGTTCAATTATTT 2449
QY 2581 ctttgtatgtatatgacacataataaatttcaatltgagaatttctcaatgtatctag 2640
|||||
Db 2450 CTPTTGATTGATATGACACATATATAAAATTTTCATTGAGAAATTTCTCAATGTATCTAG 2509
QY 2641 ttaataagcacaglttggaaactgtctgagactgacttatacaataatcaaccgacaa 2700
|||||
Db 2510 TTAATAAGCACAGTTTGGAACTTGTCTGAGACGTGACTTTATCAATAATCAACCGACAA 2569
QY 2701 agatcatatccatgtgtatggtgtagacatttcaatltgactaacccagagcagt 2760
|||||
Db 2570 AGATCATATCCATGTGATGTGGTTAGACATTTTATTTCATTGACTAACCAGGACAGT 2629
QY 2761 ttcagtgatgcaaatltgtgtgccctctgttccagctgaaacagtcctgtgactttcaaaa 2820
|||||
Db 2630 TTCAGTGATGCAAAATGTGTGCCCTCTGTGTTCAAGCTGAACAGAGTCCGTGACTTCAAAA 2689
QY 2821 ccttgaataagtcctcccaagltgtataaatltgacaatlttaggaattttaaacttga 2880
|||||
Db 2690 CCTTGAATTAAGTCTCCACAGTTGTATTAATTTGACAAATTTAGGAATTTTAAACTTTAGA 2749
QY 2881 tgaatcatttgttccatlttcaatltcaatlttcaatlttltgtltaatgcaacaggaacttaa 2940
|||||
Db 2750 TGATCATTTGGTTCCATTTTATTTCATTTTATTGTTAATGCAAAACGAGACTTAAA 2809
QY 2941 tgaacttgaatcctgtttaaagaatataaaacaatltgtatctatataatgct 3000
|||||
Db 2810 TGAACCTTGATCTCTGTTTAAAGATTATTAANAACATTTGTATCTATACATATGGCT 2869
QY 3001 ctgaggaacttagcttccaactaacactacaagatatgatactccatgtagtccatataaac 3060
|||||
Db 2870 CTTGAGGACTTAGCTTTTCACTACACTACAGGATATGATCTCCATGTAGTCCATATAAAC 2929
QY 3061 tgcagagtgatlttccagagtgctcgatactgttaattacaatcccatlaggctgaaaa 3120
|||||
Db 2930 TGCAGAGTGATTTTCCAGAGTGTCTGATACTGTATTAATTAATCTCCATTAAGGGCTGAAAA 2989
QY 3121 gaatgacctagcttctgtatacagctgtgtgtcttgaatgtgttactgtacacag 3180
|||||
Db 2990 GAATGACCTTACGTTTCTGTATACAGCTGTGTGCTTTGATGTGTACTGTACACAG 3049
QY 3181 aagtgtgtcactgaggtctctgcgtgtgttccgtatgtaaaacctgtgtagccctgcgagt 3240
|||||
Db 3050 AAGTGTGTGCACTGAGGCTCTGCGTGTGTCCTGATGGAATAACCTGTGAGCCCTGCGAGT 3109
QY 3241 taagtactgtctccatcaatgtttaaagctggaatttctcccatlgaatgtlaagtaa 3300
|||||

QY 3120 agaatgacctgaagttctctgtatatacagctgtgtgtcttctgtatgtgtgttactgtacaca 3179
|||||
Db 3121 AGAATGACCTAGCTTCTGTATACAGCTGTGTCTTGTGATGTGTGTACTGTACACA 3180
QY 3180 gaagtgtgtcgaactgagcctcgtcgtgtgtgtccgtatgtaaaacctgtgtagccctgcgag 3239
|||||
Db 3181 GAAGTGTGTGCACGTGAGGCTCTGCGTGTGTCCGTATGGAACCTGTAGCCCTGCGAG 3240
QY 3240 ttaagtactgctccatctcatctgttctcgcgtggaatttctcccatggaatgtaagta 3299
|||||
Db 3241 TTAAGTACTGCTTCATTCATGTGTAGCGCTGGAATTTTCTCCCATGGAATGTAAGTA 3300
QY 3300 aaacttaagtgttgcatacataaataatgtaactaaataaaaaa 3348
|||||
Db 3301 AAACCTAAGTGTGTGCATCAATAATGTAATACTAAAAA 3349

RESULT 2
AF061938 3217 bp mRNA PRI 04-MAR-1999
LOCUS Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
DEFINITION complete cds.
ACCESSION AF061938
VERSION AF061938.1 GI:4335946
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3217)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and Desgroselliers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3217)
AUTHORS Desgroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada

FEATURES
source 1. 3217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. 3217
/gene="STAU"
363. 1853
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AAD17530.1"
/db_xref="GI:4335947"
/translation="MKLGKKPMYKPVDPYSRMQSTYYNNMRGGAYPPRYFPVPPPL
LYVELSVGGQOENGKGTFOAKHDAALRILONEPLPERLEVNGRSEENLNK
SEISOVEFIALKRNLPVNFVARESGPRPHMKNFTYKVSVEFVGESEKSKISKNA
AIAVLEELKRLPLPAVERVKPRIKKTKPIVKPQTSPEYGGGINPISRAIOQAKK
EKEPEYTLTERGLPRRREFVMQKVGNNHTAEGTGNKKVAKRNAENMLEILGFKVP
OROPTKPKALKESEKTPIKKPKGKVTFEFGSGDENGTSNKEDFERMPYLSHQQLPA
GILPMVPEVAQAQVGSQGHHTKDFTRAAPNPAKATVTAMIARELLYGTSPTAETILK
NNISGHPHGPITRPSQDLISRVQGFQVEYKDFPKNNKNEFVSLINCSQPLIS
HGIGKDVESCHDMAALNLIKLSLDQOSTEMPRGNGPMSYCGRC"

BASE COUNT 889 a 796 c 742 g 790 t
ORIGIN

Query Match 91.9%; Score 3076; DB 10; Length 3217;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 3217; Conservative 0; Mismatches 0; Indels 131; Gaps 1;

QY 1 acttcctgcgggctgcggcgccctgagcgctcttcagcgcttgcggcgcgctgcgct 60
|||||
Db 1 ACTTCCTGCCGGGCTGCGGGCGCCTGAGCGCTCTCAGCGTTGCGGGCGGCTGCGCGT 60
QY 61 ctctctcgctcccgcttctcttgcacgcctcccccccgccggcgcgccgcctc 120
|||||
Db 61 CTCTCTCGGCTCCCGCTTCTTGTGACCGCTCCCGCCCGCGCGCGCGCGCGCTC 120
QY 121 ctcaacggccactccgcctcttccctcccttcgctcccttctctctcttctctt 180
|||||
Db 121 CTCACGCGCACCTCCGCTCTTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTT 180
QY 181 ctccctccctccctcccgccaccccgccaggaccgcccgggggagcagtcaggac 240
|||||
Db 181 CTTCCTTCCCTCCTCGCCGCCACCGCCAGGACCGCGCGGGGAGAGTCCGAGAC 240
QY 241 agcagccagcagcagcaggtgagtttgcctcttgcgccagcgctgagtgagtcg 300
|||||
Db 241 AGCAGC-----CAGAGTTTATTAACCACTTAACCTCTCAGAACTGAACAAAGAC 246
QY 301 gtgatctcgctcaactgcaacctccacctccaggtcagcgatlttccacttcagcctc 360
Db 247 ----- 246
QY 361 ccgataagctgagattacagagttatlaaccaactlaacctcagaactgaaacaagac 420
|||||
Db 247 -----CAGAGTTTATTAACCACTTAACCTCTCAGAACTGAACAAAGAC 289
QY 421 aaactgttccctggaacgcccctctttaaanaaagaagacataaccctactgtagaact 480
|||||
Db 290 AACATTGTTCTCGTAACGCCCTCTTTTAAANAAGAAAGCATAACCCCTACTGTAGAACT 349
QY 481 aaatgcactgtgcatagaacttggaaaaaaccaatgtataagcctgtgacccttact 540
|||||
Db 350 AAATGCACGTGTGCATGAACCTTGGAANAACCAATGTATTAAGCCTGTGACCCCTTACTTC 409
QY 541 tcgatgcagtcacactataactacaacatgagagagtgagtgcttaccctccgagtgact 600
|||||
Db 410 TCGGATGCAGTCCACCTATACTACAACATGAGAGAGGTGCTTATCCCGGAGGTACTT 469
QY 601 ttaccacttccagttccaccttacttactatacaagtggaacttctgtggaagacagca 660
|||||
Db 470 TTACCCATTTCAGTTCCACCTTACTTATTAAGTGAACCTTCTGTGGAGACAGACA 529
QY 661 attaatgycaaagaaagacaagacagcgtcgaacacagatgctgcgcaaacgct 720
|||||
Db 530 ATTTAATGGCAAGGAAGACAAGACAGACAGCGTCCGAAACACAGATGCTGTGCCAAAGCGTT 589
QY 721 gaggatccctgcagaatgagccctcgcagagagcgtgaggtgaatggaagaaatccga 780
|||||
Db 590 GAGGATCCTGCAGATGAGCCCCCTGCCAGAGAGGCTGAGGTGAATGGAAGAAATCCGA 649
QY 781 agaagaaatctcaataaatctgaaataagtcagtgcttgaagatgacacttaacggaa 840
|||||
Db 650 AGAAGAAATCTCAATAATCTGAATAATAGTCAAGTGTGAGATGGCACTTAACCGGAA 709
QY 841 ctgcctgtgaatttcgaaggtgccccgggagagtgccccccccacatgaagaacttgt 900
|||||
Db 710 CTTCCTGTGAATTTGAGGTTGGCCCGGAGAGTGGCCACCCACATGAAGAACCTTTGT 769
QY 901 gaccaaagtttcggttgggaggttctgtgggggaaagtgaaagggaagaagaatctc 960
|||||
Db 770 GACCAAGGTTTGGTTGGGAGTTTGTGGGGGAAAGGTGAAGGGGAAAGCAAGAAGATTTC 829
QY 961 aaagaaaatgcccagatagctgttcttgaagagctgaagaagttaccgcccctgcctgc 1020
|||||
Db 830 AAAGAAAATGCGCCCATAGCTGTCTTGAAGAGACTGAAGAAGTTACCGCCCTGCTGC 889
QY 1021 agttgaacgagtaagcctagaaatcaaaaagaanaaaccatagtcgaagccacagac 1080
|||||
Db 890 AGTTGAACGAGTAAGCCTAGATATCAAAAAGAAAACAAACCCCATTAAGCCACAGAC 949

Db 901 TGACCAAGGTTTCGGTTGGGGAGTTTGTGGGGAGAAGGTGAAGGGAAGCAAGAAGATT 960
QY 960 caaagaaaaatgcccgcataagctgttctcttgagagctgaagaagttaacgccccctgctg 1019
Db 961 CAAGAAAAATGCCGCCATAGCTGTCTTGAGGAGCTGAAGAAGTTACCGCCCCCTGCCTG 1020
QY 1020 cagttgaacgagtaaaagcctagaatcaaaaaaagaaacaaacccatagtcaagccacaga 1079
Db 1021 CAGTTGAACGAGTAAGCCTTAGAATCAAAAAAGAAACAAAACCATAGTCAAGCCACAGA 1080
QY 1080 caagcccaagaatatgcccagggygataatccgaattagccgaactgycacagatccagcag 1139
Db 1081 CAAGCCCAGAATATGGCCAGGGGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGG 1140
QY 1140 caaaaaagagaagagccagagtaacagctcctcaacagagcgagcgctcccgcgcgca 1199
Db 1141 CAAAAAAGAGAAGAGCCAGAGTACACGCTCCTCACAGAGCGAGGCTCCCGCGCGCA 1200
QY 1200 gggagtttgatgcaagtgaaagttgaaacacacactgcagaagaaagcgaccaca 1259
Db 1201 GGGAGTTTGTGATGCGAGTGAAGGTTGGAACCACTGCAGAGAAGACGGGCACCAACA 1260
QY 1260 aagaagtgcccaagcgcaatgcaagccgagaacaatgctgagatccctggtttcaaatcc 1319
Db 1261 AGAAGGTGGCCAAAGCCCAATGCAGCCGAGAACATGCTGGAGATCCTTGTTTCAAGTCC 1320
QY 1320 cgcaagcgcaagcccaacaaacccgcactcaagtcaagaggaagaacacccaataaagaac 1379
Db 1321 CGCAGCGGCGAGCCCAACAAACCCGCACCTCAAGTCAAGAGAGAAGACACCCATAAAGAAC 1380
QY 1380 cagggagtggaagaaagtaacctttttgaacctgctctggygataaagtgaacta 1439
Db 1381 CAGGGATGGAAGAAAGTAACCTTTTGAACCTGGCTCTGGGATGAAAAATGGGACTA 1440
QY 1440 gtaataaagagatgagttcaagatgacctatctaagtcatcagcagctgcctgctgaa 1499
Db 1441 GTAATAAAGAGATGAGTTCAGGATGCCTTAATCTAAGTCATCAGCAGCTGCCTGTGAA 1500
QY 1500 ttcttcccatggtgccgaggtcgcccaggtgtagagttagtcaaggaatcacacca 1559
Db 1501 TTCTTCCCATGCTGCCAGAGTCCGCCAGGCTGTAGAGTTAGTCAAGGACATCACACCA 1560
QY 1560 aagatlttacagagcagctccgaatccctgccaagggccacgtaactgccaatgataccc 1619
Db 1561 AAGATTTTACCAGGGGCAAGCTCCGAATCCTGCCAAGGCCACGGTAACCTGATAGCCC 1620
QY 1620 gagagttgtgtatggyggcaacctcgccccacagccgagaacatlttaagaataacatct 1679
Db 1621 GAGAGTTGTTGATGGGGGCACCTCGGCCACAGCCGAGACCATTTTAAGAATTAACATCT 1680
QY 1680 cttcagggcacgtaccccatgagacctctcacgagaccctctgagcaactggaatacttt 1739
Db 1681 CTTCAGGCCACGTACCCCATGAGCCTCTCAGCAGACCCTCTGAGCAACTGGACTATCTTT 1740
QY 1740 ccaagatccagagatccaagtttgaatacaaaagactccccaaaaaacaagaacgaat 1799
Db 1741 CCAGAGTCCAGGAGATCCAGGTTGAATACAAGACTTCCCCAAAAACAAGAAGCAAT 1800
QY 1800 ttgtatctctatcaaatgtctcctctcagccaacctgtatcagccatggtatcggaag 1859
Db 1801 TTGTATCTCTTATCAATTGCTCTCTCAGGCCACTCTGATCAGGCCATGGTATCGGCAAG 1860
QY 1860 atgtgagtgctcgtcagatagatgctgcgtgaaacatctlaaagtgtgctgtcagttg 1919
Db 1861 ATGTGAGTCTGCAATGATATGGCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGG 1920
QY 1920 accaacaagtacagagatgccaagaaggaacggaaccaatgtctgtgtgagaggt 1979
Db 1921 ACCAACAAGTACAGAGATGCCAAGAACAGAAACGGAACCAATGTCTGTGTGGAGGT 1980
QY 1980 gctgaacctttctcgtcattgaaccattataaatacccaacatatatactgaaaatactg 2039

Db 1981 GCTGAACCTTTTCTGGCCATGAACCATTTATAAAATCCCAACATATATACTGAANAATACTG 2040
QY 2040 aaactgctltgaaaatltgaaatltctgatacctccagltggyccgagagacaagtggt 2099
Db 2041 AAACGTCTTTGAAAAATTGGAATTTCTGATACCTCCAGTGGCGGAGAGACACGGTGGT 2100
QY 2100 aaaagatgtyggcagcagcaggygaagacaacagaaacacaaggaagcggtgtgycgcgc 2159
Db 2101 AAAGGATGTGGGAGAGAGGGAAGACAAACAGAAACACAAGGAGCGGCTGTGGCCGGC 2160
QY 2160 tggactgtcgtggygtttgtgtgatgcccacatcggtgacctgacccctacgcaata 2219
Db 2161 TGGAATGTGCTGGGTTTGTGTGATGGCCACATCGGTGACCTGGCGGTCCCTACGCAATA 2220
QY 2220 gcagctgcctgtggygaagaaaggyctgcccagccagctggtctcccggaacaccagcag 2279
Db 2221 GCAGCTGCCTGTGGGGAAGAGGGGCTGCCACAGCAGCTGGTCTCCCGGACACACGACAG 2280
QY 2280 atccacaacctggygcacctccgtgttgtgtctttttccctgtgtgaaagaagaac 2339
Db 2281 ATCCACACCCCTGGGCACCTCCGTGTTGTGCTTTTTCCTCCCTGTGTGAAGAAGAAAC 2340
QY 2340 ggcacgacccctctcaagctggtcactcagacacatlggacaacacctgacagcca 2399
Db 2341 GGCACGACCCCTTCTCAAGCTGGCTCAGCTCAGACACATTTGGACAAACCTTGACAGCCA 2400
QY 2400 tgcacagagagagcctttgaccggycccccagagctaaagcaccagagaaatcaaatgct 2459
Db 2401 TGCCAGAGAGAGGCGCTTGACCGGGCCCCAGAGCTAAAGACACAGAAATCAATGCT 2460
QY 2460 tctactcagcgtgacccaactttctagtgtgccaagggcccaacctctctgcaatc 2519
Db 2461 TCCTACTCAGCGTGACCCAACCTTTCTAGTGTGCCACAGGCCACCACTCCTGCAGTAC 2520
QY 2520 ccaacacatcaccacctgtctctctccaaacagtgatctgtatctctcaagtttcatatt 2579
Db 2521 CCACACCATCACCACCTGCTTCTCTTCCAACAGTGATCTGTATTCTTATTATTATT 2580
QY 2580 tctttgattgatatgacaactataaaatlttcaattggaatltctcaattgtatcta 2639
Db 2581 TCTTTGATTGATGACACTATATATAAAATTTCAATTTGAGAAATTTCAATTTGATCTA 2640
QY 2640 gtaaatagcacagtttggaaactgtctgagacgtacttatacaataactaacgcaga 2699
Db 2641 GTTAAATAGCACAGTTTGGAAACTTGTCTGAGACTGACTTATCAATTAATCTAACCGACA 2700
QY 2700 aagatcatatccatgtgtatgtgttagacatltttatctatctgactaaacccagagcag 2759
Db 2701 AAGATCATATCCATGTGTATGTGTTAGACATTTTATTTCAATTGACTTAACCCAGACAG 2760
QY 2760 ttccagtgatgcaaatgtgtgcccctcgtgttcagctgaaacagctccctggaactttcaaa 2819
Db 2761 TTTTCAGTATGCAAAATTTGTGTGCCCTCTGTGTTCAAGCTGAACAGTCCCTGACTTTCAAAA 2820
QY 2820 acctgaaataagctcccaacagltgtataaattggaacaatttaggaatlttaaccltag 2879
Db 2821 ACCTGAATTAAGTCTCCACAGTTGTATTAATTTGACAAATTTAGGAATTTTAAACTTTAG 2880
QY 2880 atgatacttggttccatltttatctatcttcaatttltgltaatgcaaacaggaactaa 2939
Db 2881 ATGATCATTTGTTCCATTTTATTTTCAATTTTATTTTGTTAATGCAACAGACTTAA 2940
QY 2940 atgaactttgatctcgttttaagaattataaaaaacatgtgtatcctatacatatgyc 2999
Db 2941 ATGAACCTTGATCTCTGTTTAAAGATTAATTAACAAACATTTGTATCTATATATGCGC 3000
QY 3000 tcttgaggaactagcttcaactacactacagagatatgatactccatgtagttccatataaac 3059
Db 3001 TCTTGAGGACTTAGCTTTCACCTACACTACAGGATATGATCTCCATGTAGTCCATATAAAC 3060
QY 3060 ctgcagagtgatlttccagagtgctcgatactgttlaattacatctccatlagggctgaaa 3119
Db 3061 CTGCAGAGTGATTTTCCAGAGTGTCTGATACTGTTAATTACATCTCCATTTAGGGCTGAAA 3120

22	150.8	4.5	160979	52	AC055879	AC055879	Homo sapi
23	149.6	4.5	3173	34	AK001576	AK001576	Homo sapi
24	149	4.5	332811	72	AL390198	AL390198	Homo sapi
25	133.4	4.0	194407	54	AC068845	AC068845	Homo sapi
26	125.2	3.7	2008	34	AK023314	AK023314	Homo sapi
27	119	3.6	2847	69	HSRNATFTFI	X83973 H.sapiens m	
28	110.2	3.3	5237	53	DROSTAUFEN	M69111 D.melanogas	
29	110.2	3.3	61267	28	AC004336	AC004336	Drosophill
30	110.2	3.3	120625	38	AC017563	AC017563	Drosophill
31	110.2	3.3	267271	28	AE003800	AE003800	Drosophill
32	105.8	3.2	8409	29	AF225924	AF225924	Drosophill
33	103	3.1	8002	29	AF225925	AF225925	Musca dom
34	98.6	2.9	109088	34	AL137011	AL137011	Human DNA
35	98.2	2.9	197385	37	AC015936	AC015936	Homo sapi
36	96	2.9	148030	63	AL137803	AL137803	Homo sapi
37	95	2.8	170559	37	AC016253	AC016253	Homo sapi
38	94.8	2.8	216407	27	AC009660	AC009660	Homo sapi
39	94.6	2.8	149711	49	AC025869	AC025869	Homo sapi
40	94.2	2.8	41488	26	AC006180	AC006180	Homo sapi
41	94.2	2.8	131518	56	AC074090	AC074090	Homo sapi
42	94.2	2.8	150834	52	AC063960	AC063960	Homo sapi
43	93.6	2.8	167335	48	AC023527	AC023527	Homo sapi
44	93.6	2.8	177902	49	AC025936	AC025936	Homo sapi
45	93.6	2.8	197778	10	AF130358	AF130358	Homo sapi

ALIGNMENTS

```

RESULT      1
AF061941
LOCUS
DEFINITION  AF061941      3349 bp      mRNA      PRI      04-MAR-1999
Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION   AF061941
VERSION     AF061941.1  GI:4335952
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3349)
AUTHORS     Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE       Mammalian stauflen is a double-stranded RNA and tubulin binding
            protein which localizes to the rough endoplasmic reticulum
            Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL     99147057
MEDLINE
REFERENCE   2 (bases 1 to 3349)
AUTHORS     DesGroselliers,L., Wickham,L. and Luo,M.
TITLE       Direct Submission
JOURNAL     Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
            Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
            source
                1..3349
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="20"
                /map="20q13.1"
                1..3349
                /gene="STAU"
                495..1985
                /gene="STAU"
                /function="RNA binding protein"
                /note="alternatively spliced transcript"
                /codon_start=1
                /product="stauflen protein"
                /protein_id="AAD17533.1"
                /db_xref="GI:4335953"
                /translation="MKIGKKPMYKPPVDYSRMSTYNNMRGAYPPRYFPFPVPL
                LYQVELSVGGOQFNGKGTROAAKHDAAKALRILQNEPLPERLEVNGRESEENLNK
                SEISOVEIALKRNLPVNFEEVARESGPHMKNEVTYVSGEFGEGSKSKISKNA
                AIAVLEELKKLPLPAVERVKPRIKKTPIVKPQTSPEYGGINPISRLAQIQAKK
                EKEPEYTLTLTERGLPRRREFVMOVKVGNTAEGTGNKKVAKRNAEENMLEILGFKVP

```

BASE COUNT 913 a 838 c 777 g 821 t

ORIGIN

QROPTKPKLSEKPTIKKPGDGRKVTFFEEBGSDENGTSNKKEDEFRMPLYSHOOLPA
 GILPMVEVAQAVGSQGHHTKDFTRAPNPAKATVTAMIARELLYGTSPTAETILK
 NNISGHPHGPLTRPSEOLDYLSRVQGEYEEKDEPKNNKNEFVSLNCSOPLISS
 HGIGKDVESCHDMAALNIIKLSELDOOSTEMPRGTNGPMSSVCGRC"

Query Match	99.7%;	Score 3337;	DB 10;	Length 3349;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3348;	Conservative	0;	Mismatches	0;
			Indels	1;
			Gaps	1;

[illegible]


```

misc_feature      /note="assembly_fragment"
67108. .76018
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
76119. .85029
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
85130. .92117
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
92218. .103464
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
103565. .111248
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
111349. .122466
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
122567. .132062
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
132163. .148654
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
148755. .160979
/note="assembly_fragment
clone_end:T7
vector_side:left"
BASE COUNT      44926 a 35777 c 35800 g 42072 t 2404 others
ORIGIN

```

Query Match	8.3%;	Score 291;	DB 52;	Length 160979;
Best Local Similarity	100.0%;	Pred. No. 3.1e-57;		
Matches 291; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	323	aggtagaactcttagaacttcatagacagtgaattaaacctgcactyaaaagctgtttactcyca	382
Db	11271	AGGTAGAACCTTTAGACTTTCATAGCAGCTGAATTAACTGCACTGAAAAGCTGTTTACCCTGCA	11212
OY	383	tltgttcaacttttgttgaaagtgaccatgltcctaagtlccaagtycaaagtlcagaaccat	442
Db	11211	TTTGTCACCTTTTGTTGAAAGTGACCATGTCTCAAGTTCACAGTGCAAAGTTCAGAACCCAT	11152
OY	443	ctgctgctctctcagggagccaaatactgaacaagaaccagtcctctctctcacagcctt	502
Db	11151	CTGCTGCTCTCTCAGGGAGGCCAAATACTGAACAAGAACCAGTCTTCTCTCACAGCCTT	11092
OY	503	tgatgagtatctccctctactactagctctctgcgccctctgaaaaatgcaagttagaccatctc	562
Db	11091	TGATGAGTAGTATCTCTTCTACTACTAGCTCTTGCCCCCTGGAANAATGCAGGTAGACCCCATTC	11032
OY	563	aaaactctgctttaaaccctctgcatactaatcatccaaccaagtycagctycag	613
Db	11031	AAAACTCTGCTTTTAACTCCCTCTGCATCTATTACATCCACCACAGTGCAGCTGCAG	10981

Search completed: April 5, 2001, 09:07:34
Job time: 79699 sec

AUTHORS
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivar, T.M., Oliver, J., Peterson, K., Plerre, N., Pisaní, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT

Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 22, 2000 this sequence version replaced gi:7582713.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997) \
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L9868
 Center clone name: 746_F_15

----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 147035 bases at least Q40
 Consensus quality: 153931 bases at least Q30
 Consensus quality: 156692 bases at least Q20
 Insert size: 167000; agarose-fp
 Insert size: 158579; sum-of-contigs
 Quality coverage: 4.0 in Q20 bases; agarose-fp
 Quality coverage: 4.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1728:	contig of 1728 bp	in length
*	1729	1828:	gap of 100 bp
*	1829	3355:	contig of 1527 bp in length
*	3356	3455:	gap of 100 bp
*	3456	6090:	contig of 2635 bp in length
*	6091	6190:	gap of 100 bp
*	6191	9106:	contig of 2916 bp in length
*	9107	9206:	gap of 100 bp
*	9207	12073:	contig of 2867 bp in length
*	12074	12173:	gap of 100 bp
*	12174	16121:	contig of 3948 bp in length
*	16122	16221:	gap of 100 bp
*	16222	19552:	contig of 3331 bp in length
*	19553	19652:	gap of 100 bp
*	19653	23309:	contig of 3657 bp in length
*	23310	23409:	gap of 100 bp
*	23410	27884:	contig of 4475 bp in length

*	27885	27984:	gap of	100 bp	
*	27985	32689:	contig of	4705 bp	in length
*	32690	32789:	gap of	100 bp	
*	32790	36488:	contig of	3699 bp	in length
*	36489	36588:	gap of	100 bp	
*	36589	43152:	contig of	6564 bp	in length
*	43153	43252:	gap of	100 bp	
*	43253	46195:	contig of	2943 bp	in length
*	46196	46295:	gap of	100 bp	
*	46296	51596:	contig of	5301 bp	in length
*	51597	51696:	gap of	100 bp	
*	51697	56969:	contig of	5273 bp	in length
*	56970	57069:	gap of	100 bp	
*	57070	67007:	contig of	9938 bp	in length
*	67008	67107:	gap of	100 bp	
*	67108	76018:	contig of	8911 bp	in length
*	76019	76118:	gap of	100 bp	
*	76119	85029:	contig of	8911 bp	in length
*	85030	85129:	gap of	100 bp	
*	85130	92117:	contig of	6988 bp	in length
*	92118	92217:	gap of	100 bp	
*	92218	103464:	contig of	11247 bp	in length
*	103465	103564:	gap of	100 bp	
*	103565	111248:	contig of	7684 bp	in length
*	111249	111348:	gap of	100 bp	
*	111349	122466:	contig of	11118 bp	in length
*	122467	122566:	gap of	100 bp	
*	122567	132062:	contig of	9496 bp	in length
*	132063	132162:	gap of	100 bp	
*	132163	148654:	contig of	16492 bp	in length
*	148655	148754:	gap of	100 bp	
*	148755	160979:	contig of	12225 bp	in length

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-746F15"
/clone_lib="RPC1-11 Human
1.1728
misc_feature
/note="assembly_fragment"
1829..3355
/note="assembly_fragment"
3456..6090
/note="assembly_fragment"
6191..9106
/note="assembly_fragment"
9207..12073
/note="assembly_fragment"
12174..16121
/note="assembly_fragment"
16222..19552
/note="assembly_fragment"
19653..23309
/note="assembly_fragment"
23410..27884
/note="assembly_fragment"
27985..32689
/note="assembly_fragment"
32790..36488
/note="assembly_fragment"
36589..43152
/note="assembly_fragment"
43253..46195
/note="assembly_fragment
clone_end:SP6
vector_side:right"
46296..51596
/note="assembly_fragment"
51697..56969
/note="assembly_fragment"
57070..67007

```


Db 60502 ACATCTTTTCTTACTGGACTAACCAGAACCATTTTCAGTGATGCAATCATGTGACCTCT 60443
QY 2946 ggttcagctgaacacagtcctcgtgaact-tcaaaaaccttgaataagtcctccacagttgta 3004
Db 60442 GGTTCAGCTGAAAAAGCCCTGGACCTTCTCAGAAACCTTGATCAAAATTTTCCACTATTTGTA 60383
QY 3005 taattgacaatttaggaatttaaaccttagatgatcatctgttcacatttatttc 3064
Db 60382 CAAATGGCCATTTAGGAATTTCAACCTCA---ATCATTTGGCTCTCTTTATATTTT 60327
QY 3065 atttatttttgtaatagca--aacaggaacttaaatgaactttgatctctgttttaag 3122
Db 60326 ATTTTATTTTGGTTGTGTAATGGAGCCTAACAATAAATTAACCTTGATCTTTGTTTAAAG 60267
QY 3123 attatataaaaacattgtgtatctatacatatagctccttgaggacttagctttcactaca 3182
Db 60266 ATTATTAATAAACCTTACGCTGATGTACATATAGCTCTTCAAGACCCACAGCTTTCACCTATG 60207
QY 3183 ctacagatatatctccatcgttagtcacataaaacctgcagagtgaatttcc-agaatgc 3241
Db 60206 CTGTAGGATATGTCTCTCTGTAGTAGACATATGACATTTGACAGGTTAATTTTCCCTTGAGTGC 60147
QY 3242 tcgatactgttaattacatc 3261
Db 60146 TTTTACTCTTATTTTCTCTC 60127

RESULT 14
G30939/c 385 bp DNA STS 29-SEP-1998
LOCUS G30939 385 bp DNA STS 29-SEP-1998
DEFINITION DRES50 Human DRES Homo sapiens STS genomic clone 22368 3', sequence
tagged site.
ACCESSION G30939
VERSION G30939.1 GI:1724032
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 385)
AUTHORS Banfi,S., Borsani,G., Rossi,E., Bernard,L., Guffanti,A.,
Rubboli,F., Marchitelli,A., Giglio,S., Coluccia,E., Zollo,M.,
Zuffardi,O. and Ballabio,A.
IDENTIFICATION AND MAPPING human cDNAs homologous to Drosophila
mutant genes through EST database searching
JOURNAL Nat. Genet. 13 (2), 167-174 (1996)
MEDLINE 96225443
COMMENT

Contact: Ballabio Andrea
Telethon Institute of Genetics and Medicine, TIGEM
Telethon Institute of Genetics and Medicine, TIGEM
via Olgettina 58, 20132 Milano, Italy
Tel: +39-2-215601
Fax: +39-2-21560220
Email: dres@tigem.it
Primer A: GTAACAATGATGGAACGAG
Primer B: GGCGTGAAGAAGATGACCTAC
STS size: 157
PCR Profile:

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 220 uM
Taq Polymerase: 0.025 units/ul
Total Vol: 15 ul
Presoak: 94 degrees C for 2 minutes
Denaturation: 94 degrees C for 45 seconds
Annealing: 57 degrees C for 45 seconds
Polymerization: 72 degrees C for 45 seconds
PCR Cycles: 35
Thermal Cycler: MJR PTC-200

Buffer: MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
The 5' end of the cDNA clone 22368 is homologous to the staufen
Drosophila gene product. EST from which STS is derived: T89105.
Forward primer is TO-1233; Reverse primer is TO-1234.
FEATURES
source
1. .385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="20, 12.10 cR from GCT10F11 (lod >3.0)"
/clone="22368"
STS
primer_bind 29. .185
primer_bind 29. .49
BASE COUNT 135 a 79 c 71 g 100 t
ORIGIN

Query Match 10.8%; Score 380.2; DB 75; Length 385;
Best Local Similarity 99.2%; Pred. No. 2.1e-78;
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3069 ttattttgttaatgaacaggaacttaatgaactttgatctctgttttaagattatt 3128
Db 385 TTTATTTTGTATGCAACAGGACTTAATGAACCTTGATCTCTGTTTAAAGATTATT 326
QY 3129 aaaaacattgtgtatctatacatatgctccttgaggacttagctttcactacacag 3188
Db 325 AAAAAACATGTGTATCTATACATATGCGCTTGAGGACTTAGCTTTCACACTACAG 266
QY 3189 gatatgctcctcatgtagtcacataaaacctgcagagtgaatttccagagtgctgatac 3248
Db 265 GATATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGATTTTCCAGAGTGCTGATAC 206
QY 3249 tgttaattacatctccattagggctgaaagaatgaacctacgtttctgtatcacgctgtg 3308
Db 205 TGTTAATTACATCTCCATTTAGGGCTGAAAGAATGACCTACGTTTCTGTATACAGCTGTG 146
QY 3309 ttgctttgatgtgtgtactgttacacagaagtgtgtgcaactgaggtctgcgtgtgt 3368
Db 145 TTGCTTTTGATGTGTGTACTGTACACAGAGTGTGTGACACTGAGGCTCTGCGTGTGT 86
QY 3369 ccgtatgaaacacctgtagccctgcgagtgtaagtactgcttccattcatgtttaagct 3428
Db 85 CCGTATGGAAGCCCTGGTAGCCCTGCGAGTTAGTACTGCTTCATTCATTGTTACCGT 26
QY 3429 ggaattttctcccatggaatgta 3453
Db 25 GGAATTTTCTCCCATGGAATGTA 1

RESULT 15
AC055879/c 160979 bp DNA HTG 22-MAY-2000
LOCUS AC055879 160979 bp DNA HTG 22-MAY-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-746F15 map 15, WORKING DRAFT
SEQUENCE, 25 unordered pieces.
ACCESSION AC055879
KEYWORD AC055879.2 GI:8017940
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 160979)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-746F15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160979)

RESULT	13
LOCUS	AL158850/c
DEFINITION	AL158850 192597 bp DNA HTG 06-SEP-2000 Homo sapiens chromosome 6 clone RP3-522B19, *** SEQUENCING IN PROGRESS ***, 11 unordered pieces.
ACCESSION	AL158850
VERSION	AL158850.4 GI:9800143
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 192597) Burton,J.
REFERENCE	Direct Submission Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Aug 13, 2000 this sequence version replaced qi:9212635.
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: d0522B19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 12% of reads
Sequencing vector: plasmid; L08752; 87% of reads
Chemistry: Dye-terminator Big Dye; 77% of reads
Chemistry: Dye-primer-amersham; 22% of reads
Consensus quality: 185625 bases at least Q40
Consensus quality: 188533 bases at least Q30
Consensus quality: 190005 bases at least Q20
Insert size: 191597; sum-of-contigs
Insert size: 144116; 9.6% error; agarose-fp
Quality coverage: 5.71x in Q20 bases; sum-of-contigs Quality
coverage: 8.21x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 13837: contig of 13837 bp in length
* 13838 13937: gap of 100 bp
* 13938 17476: contig of 3539 bp in length
* 17477 17576: gap of 100 bp
* 17577 20005: contig of 2429 bp in length
* 20006 20105: gap of 100 bp
* 20106 32070: contig of 11965 bp in length
* 32071 32170: gap of 100 bp
* 32171 39772: contig of 7602 bp in length
* 39773 39872: gap of 100 bp
* 39873 43471: contig of 3599 bp in length
* 43472 43571: gap of 100 bp
* 43572 51795: contig of 8224 bp in length
* 51796 51895: gap of 100 bp
* 51896 178838: contig of 126943 bp in length
* 178839 178938: gap of 100 bp
* 178939 181919: contig of 2981 bp in length
* 181920 182019: gap of 100 bp
* 182020 184231: contig of 2212 bp in length
* 184232 184331: gap of 100 bp
* 184332 192597: contig of 8266 bp in length.
Location/Qualifiers

```

```

source
1. .192597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-522B19"
/clone_lib="RPC1-3"
1. .13837
/note="assembly_fragment:02383
fragment_chain:1"
13938. .17476
/note="assembly_fragment:01080
fragment_chain:1"
17577. .20005
/note="assembly_fragment:02434
fragment_chain:1"
20106. .32070
/note="assembly_fragment:03060
fragment_chain:2"
32171. .39772
/note="assembly_fragment:00945
fragment_chain:2"
39873. .43471
/note="assembly_fragment:00013
fragment_chain:2"
43572. .51795
/note="assembly_fragment:02225
fragment_chain:3"
51896. .178838
/note="assembly_fragment:01767
fragment_chain:3"
178939. .181919
/note="assembly_fragment:01637"
182020. .184231
/note="assembly_fragment:02408"
184332. .192597
/note="assembly_fragment:02960"
39515 c 38313 g 55431 t 1014 others
BASE COUNT 58324 a
ORIGIN

```

Query Match	10.9%;	Score 383.2;	DB 70;	Length 192597;
Best Local Similarity	73.2%;	Pred. No. 8e-79;		
Matches 586;	Conservative 0;	Mismatches 203;	Indels 11;	Gaps 7;
QY 2466	gltctttttttccctgtgtgaaagaagaacagcgacgacccttcctcaagctgctcac	2525		
Db 60919	gTATTTTATTTGGCTGTGAAANAAGAAATGGAAGACAGAAAGCATTAACCTGCTCAC	60860		
QY 2526	tcagacacatttgggacaaaaccttggacagccatgccaagagagagccttgcacggccc	2585		
Db 60859	ACTGGCTGACTCAAAC - ACCTTGCTATAAACATGCTGGAGAAAAGCCTTAGACTGGCCTC	60801		
QY 2586	agagctaaagcaccagagaaatcaaatgcttcctactcagcgtgaacccaactttcta	2645		
Db 60800	AAAGCTGAAAACACCGGGGAAATTAACCTGCTTCTACTCGTTACACATTACCTTCCTA	60741		
QY 2646	gtgtgccagggcccaaccactctgcagtlaccacacacatcaccactgtctctcc	2705		
Db 60740	GCATGCTACGG - CCAGTCACCTTACTGCAGTACCACACCATTAACCACTGCTTCTTCT	60682		
QY 2706	aacagtgatctgtatctctagtcttcattatcttcttggatgtatgacactatataa	2765		
Db 60681	AATAGTGATACATATTTTACTTTATTA - TTTCTTTGATTGAATGATGCCATATGAA	60623		
QY 2766	atttcatlttggaatttctcaattgtatctagttaatatagcacagtttggaaactgtc	2825		
Db 60622	ATTTCATTGGAAGTTTCTCAGTTGTATCTAGTTACATAGTACAGTTTCAGAAACCTGTC	60563		
QY 2826	tgagactgacttatcaataatcctaaccgacaaagatcatatccatgtgtatgttag	2885		
Db 60562	CCAGACTGACTTTATTCAGGAATCTAACGTGCATAGATCATATTCATGTATACGTGTTGT	60503		
QY 2886	acattttattcatctgactaacccaggaacglttcagltgataaattgtgtccctct	2945		

*	53039	59129: contig of 6091 bp in length
*	59130	59229: gap of 100 bp
*	59230	65896: contig of 6667 bp in length
*	65927	65996: gap of 100 bp
*	65997	89427: contig of 2331 bp in length
*	89428	89527: gap of 100 bp
*	89528	95785: contig of 6258 bp in length
*	95786	95885: gap of 100 bp
*	95886	102762: contig of 6877 bp in length
*	102763	102862: gap of 100 bp
*	102863	110899: contig of 8037 bp in length
*	110900	110999: gap of 100 bp
*	111000	121002: contig of 10003 bp in length
*	121003	121102: gap of 100 bp
*	121103	133048: contig of 11946 bp in length
*	133049	133148: gap of 100 bp
*	133149	145755: contig of 12607 bp in length
*	145756	145855: gap of 100 bp
*	145856	165006: contig of 19151 bp in length
*	165007	165106: gap of 100 bp
*	165107	188243: contig of 23137 bp in length
*	188244	188343: gap of 100 bp
*	188344	194407: contig of 6064 bp in length.

FEATURES	Location/Qualifiers
source	1. .194407

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19"
/clone="RP11-73711"
/clone_lib="RPC1-11 Human Male BAC"
1. .1755
/misc_feature
/feature="assembly_fragment"
1856. .3429
/feature="assembly_fragment"
3530. .5346
/feature="assembly_fragment"
5447. .7338
/feature="assembly_fragment"
7439. .10297
/feature="assembly_fragment"
10398. .13590
/feature="assembly_fragment"
13691. .16293
/feature="assembly_fragment"
16394. .19065
/feature="assembly_fragment"
19166. .22713
/feature="assembly_fragment"
22814. .27165
/feature="assembly_fragment"
27266. .32377
/feature="assembly_fragment"
32478. .36572
/feature="assembly_fragment"
36673. .41657
/feature="assembly_fragment"
41758. .47024
/feature="assembly_fragment"
47125. .52938
/feature="assembly_fragment"
53039. .59129
/feature="assembly_fragment"
59230. .65896
/feature="assembly_fragment"
65997. .89427
/feature="assembly_fragment"
89528. .95785
/feature="assembly_fragment"
95886. .102762
/feature="assembly_fragment"
102663. .110899
/feature="assembly_fragment"

```

misc_feature	111000. .121002
	/note="assembly_fragment"
misc_feature	121103. .133048
	/note="assembly_fragment"
misc_feature	133149. .145755
	/note="assembly_fragment"
misc_feature	145856. .165006
	/note="assembly_fragment"
misc_feature	165107. .188243
	/note="assembly_fragment"
misc_feature	188344. .194407

```
vector_side:right"
```

[illegible]

ORIGIN

Query Match	15.8%;	Score 553.2;	DB 54;	Length 194407;
Best Local Similarity	90.4%;	Pred. No. 1.2e-118;		
Matches 665; Conservative	0;	Mismatches 48;	Indels 23;	Gaps 6;

QY	2041	gctgcgctgacatcttaagttgctgtctgagttgaccaaagttacagatgcca	2100
Db	3297	GCTGCACCTGAACACCTTAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCA	3238

QY 2101 agaacaggaacgacccaatgtctgtgtgtggaggtgtctgaacctttctgycatgaa 2160

Db 3237 AGACAGGAATGATCAATGCTCTGTGTGGGAGGTGCTGAACCTTTTCTGGCCATGAA 3178

QY 2161 ccattataaatcccaacatatatactgaaataactgaaactgcttgaagaatttggaaat 2220

Db 3177 CCATTATAAATCCCAACATATACACTGAAATATTTAAACTGCTTTGAAATTTGGAAT 3118

QY 2221 ttctgataccttcagtgggccgagagacacggtggtlaagatgtggcagcagcagg 2280

Db 3117 TTCTGATACCTCTAGTGGCCGAGAG--ACGGTGGGTAAGGATGTGGCAGCAGCAGGG 3060

QY 2281 aagacaacagaaacacaagagcgctgtgccc-ggctgactgtgctgggttltg 2339

Db 3059 AACACACAG-----GGAGGTGGCTGTGCGCCGGGCTGACTGTGCTGGGGTTGTTG 3007

QY 2340 tgatgccaactcggtgacctgcggtccctacgcaatagcagctgcctgtgggaagaag 2399

Db 3006 TGACGGCCACTCGGTGACCTGGCTGTCCCTACGCAATACCAAGCTGCCCTTGGGGAAGAAG 2947

QY 2400 qqctqcccaqccaqctqttctctccqgqacacccaqcgagatccacacccctqgqcaactccg 2459

Db 2946 GGCTGCCAGCCAGCTGCTCTCTGGGGCACCAGCAGATCCACGCCCTGGGCACCTCC 2887

Ov 2460 tatltgattcttttttccctctg--tgaagaagaaacacgacgaccccttctcaac 2517

Db 2886 TGTGTGGCTTTTTCCTGTGAAGAAGAACTGAAGCATGACCCCTTCTCAAGC 2827

2518 taactcactcaaacacat taagaaacaaacccctaaacaaaccat accaagaaagaccttga 2577

Dh 2825 TGGCTTA-----GTTGGGACCAACCTTGACACGCCACACAGAGAGAGGCTTCAA 2776

Ov 2578 ccgaccccaaacataaaacacaccaaagaaaaatcaaatcctcctactcaacgtgacccaa 2637

db 2775 CCAGCCCCAGAGCTAAAAAGCACCAGAGAAAGTCAATAATGCTTCTATTTCAGCATGACCTAA 2716

Ov 2638 attttctaatatccccccccccccccccctctacaaatacccccaaaccaatcaaacctactt 2697

Db 3715 C - - TTTCTACTCTGCGCCACCGCCCGCCACTTCTCTGCGCCACCGCCACTGCGTTT 2658

3698 tctcttcccaacatnctctatctcttaatttcattatctctttgaattgaatatgacac 2757

[illegible]

0258 ++++++Cat 0773 .

Db 3587 MATATAAATTTCCT 3583

Db 149889 TGCTGGGATTACTGGCATGAGCCACCGTGCCCGAGACACTATATAAATTTTCATTGCA 149830

Oy 2778 gaatttccaatttgtatctagttaaatagcacagtttggaaaacttgtctgagactgacct 2837

Db 149829 GAGTTTCTCAACTGTATCTAATTAATATCGCACAGTTTGGAACCTTTTCTGAGACTGACTT 149770

Oy 2838 tatcaataatctaaccgacaaagatcatatccatgtgtatgtgttagacattttatt 2897

Db 149769 TATCAATAATCTAACCAGCAAGATCATATCCACGTGTATGTAGTACATTTTATTTT 149710

Oy 2898 catgactaacccagacagtttcagtgatgcaaatgtgtgccctctgttcagctgaa 2957

Db 149709 CATTGACTAACCCAGCACCATTTCAAGTGAATTAAGTCTCCACAGTTGTATAAATTGACAAT 149650

Oy 2958 acagtcctggacttccaacaccttgataagctctccacacagttgtataaatggacaat 3017

Db 149649 ACAGTCTTGACTTTCAAAAACCTTGAATAAGTCTCCACAGTTGTATAAATTGACAAT 149590

Oy 3018 ttaggaattttaaactttagatgatcatcttggttccatttttattcattttatttg 3077

Db 149589 TTAGGAATTTTAACCTTTAGATGATCATTTGCTTCCATTTTATTTTATTTTATTTTTC 149530

Oy 3078 ttaatgcaaacagagact---aaatgaacttgatctctgtttaaagattataaaaa 3133

Db 149529 TTAATGCAAAACAGACTTAACAATGAACCTTTGATCTCCATTTTAATATATCTTTAAGA 149470

Oy 3134 acatgtgtatcatatacatatgctcttgagagacttagactttcactacacagatat 3193

Db 149469 AAATCAATGTGCTTATACATACGCTCTTGAGAGACTTGACTTTCACACTACACATACAGATAT 149410

Oy 3194 gatctccatgtagtcacataataacctgcagagtgatattccagagtgctcgatactgtta 3253

Db 149409 GATCTCCACATAGTTTACACATAAAACGACAGACTGATTTTCCAGAGTGCCTCGATACTATTA 149350

Oy 3254 attacatctccatttagggctgaaaagaatgacctacgtttctgtatatacagctgttgc 3313

Db 149349 ATTACATCTCGATTAGGGCTGAAGAGTATGACCCTATGTTTCTTTATACAGTGTGTGCT 149290

Oy 3314 ttgatgttgtgttactgtacacagaagtgtgtgcactgagc 3356

Db 149289 TTGTATGTGTGTGTACTATACAGAGAGTGTGTGACTGGGCC 149247

RESULT 12

AC068845/c

LOCUS AC068845 194407 bp DNA HTG 04-SEP-2000

DEFINITION Homo sapiens chromosome 19 clone RP11-73711 map 19, WORKING DRAFT

SEQUENCE, 27 unordered pieces.

ACCESSION AC068845

VERSION AC068845.3 GI:9966312

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 194407)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 194407)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczkyl,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., Mcpheeters,R.,

Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,

TITLE

JOURNAL

COMMENT

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 4, 2000 this sequence version replaced gi:8247880.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7549

Center clone name: 737_I_1

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 181875 bases at least Q40

Consensus quality: 187842 bases at least Q30

Consensus quality: 190220 bases at least Q20

Insert size: 190000; agarose-efp

Insert size: 191807; sum-of-ctnigs

Quality coverage: 4.7 in Q20 bases; agarose-efp

Quality coverage: 4.7 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 27 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1755: contig of 1755 bp in length

* 1756 1855: gap of 100 bp

* 1856 3429: contig of 1574 bp in length

* 3430 3529: gap of 100 bp

* 3530 5346: contig of 1817 bp in length

* 5347 5446: gap of 100 bp

* 5447 7338: contig of 1892 bp in length

* 7339 7438: gap of 100 bp

* 7439 10297: contig of 2859 bp in length

* 10298 10397: gap of 100 bp

* 10398 13590: contig of 3193 bp in length

* 13591 13690: gap of 100 bp

* 13691 16293: contig of 2603 bp in length

* 16294 16393: gap of 100 bp

* 16394 19065: contig of 2672 bp in length

* 19066 19165: gap of 100 bp

* 19166 22713: contig of 3548 bp in length

* 22714 22813: gap of 100 bp

* 22814 27165: contig of 4352 bp in length

* 27166 27265: gap of 100 bp

* 27266 32377: contig of 5112 bp in length

* 32378 32477: gap of 100 bp

* 32478 36572: contig of 4095 bp in length

* 36573 36672: gap of 100 bp

* 36673 41657: contig of 4985 bp in length

* 41658 41757: gap of 100 bp

* 41758 47024: contig of 5267 bp in length

* 47025 47124: gap of 100 bp

* 47125 52938: contig of 5814 bp in length

* 52939 53038: gap of 100 bp

AC011492/c
LOCUS AC011492 157803 bp DNA HTG 04-MAY-2000
DEFINITION Homo sapiens chromosome 19 clone CTB-187L3, WORKING DRAFT SEQUENCE,
8 unordered pieces.
AC011492
ACCESSION AC011492
VERSION AC011492.5 GI:7690131
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 157803)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157803)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 4, 2000 this sequence version replaced gi:7458736.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 136279, BC101302
Center clone name: CIT978SKB_187L3

Summary Statistics
Consensus quality: 152548 bases at least Q40
Consensus quality: 153824 bases at least Q30
Consensus quality: 154424 bases at least Q20
Estimated insert size: 156750; agarose-fp estimation
Estimated insert size: 157103; sum-of-ctigs estimation
Quality coverage: 9.52 in Q20 bases; agarose-fp estimation
Quality coverage: 9.5 in Q20 bases; sum-of-ctigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2374: contig of 2374 bp in length
* 2375 2474: gap of unknown length
* 2475 4521: contig of 2047 bp in length
* 4522 4621: gap of unknown length
* 4622 7566: contig of 2945 bp in length
* 7567 7666: gap of unknown length
* 7667 12083: contig of 4417 bp in length
* 12084 12183: gap of unknown length
* 12184 18288: contig of 6105 bp in length
* 18289 18388: gap of unknown length
* 18389 63650: contig of 45262 bp in length
* 63651 63750: gap of unknown length
* 63751 111373: contig of 47623 bp in length
* 111374 111473: gap of unknown length
* 111474 157803: contig of 46330 bp in length.

FEATURES
Source location/Qualifiers
1.157803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTB-187L3"
/clone_lib="CalTech human BAC library B"

BASE COUNT 40983 a 38390 c 39640 g 38079 t 711 others
ORIGIN

Query Match 20.1%; Score 705; DB 35; Length 157803;

Best Local Similarity 72.5%; Pred. No. 3.3e-154;
Matches 1206; Conservative 0; Mismatches 90; Indels 367; Gaps 8;

QY 2041 gctgcgtgaacatcttaaagttgctgtctgagttgaccacaagaagtacagagatgcca 2100
|||||
Db 150889 GCTGCACCTGAACACCTTAAGTTGCTGCTGAGTTGGACCAACAAGTACAGAGATGCCA 150830
QY 2101 agaacaggaacgcgaaccaatgtctgtgtgtggaagtgctgaaccccttctgccaatgaa 2160
|||||
Db 150829 AGAACAGGAATGATGATCAATGTCTGTGTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAA 150770
QY 2161 ccaataaatacccaacatatataactgaaataactgaactgcttgaanaattggaat 2220
|||||
Db 150769 CCATTATAAAATCCCAACATATACACTGAGAATATTAATAACTGCTTTGAATAATTGGAAT 150710
QY 2221 ttctgatacctccagtgggccgagagacacggtggtgaagatgtgagcagcaggg 2280
|||||
Db 150709 TTCTGATACCTTAGTGGGCCGAGAG--ACGGTGGTAAAGGATGTGGCAGCAGCAGGG 150652
QY 2281 aagacaacagaaacacaagaagagcggtgtggcc-ggctgagctgtgctgggtgtgtg 2339
|||||
Db 150651 AACACACAG-----GGGAGGTGGCTGTGGCCGGGCTGACTGTGCTGGGGTTGTTG 150599
QY 2340 tgatggcactcgtgtgacctgtgcggtccctacgcgaatagcagctgcctgtggggaagaag 2399
|||||
Db 150598 TGACGGCCACTGGGTGACCTGGCTGTCTTACGCAATACACAGCTGCCCTTTGGGGAAGAAG 150539
QY 2400 ggctgccagcagctgtgtctcccgagacaccagcagatccacccctggcactccg 2459
|||||
Db 150538 GGCTGCCAGCCAGCTGGCTCTCCTGGGGCACACAGAGATCCAGCCCTGGGACACTTCC 150479
QY 2460 tgtttgtcttttttccctgtg--tgaagaagaacgcgaccccttctcaagc 2517
|||||
Db 150478 TGTTTGCTTTTTCCTCCCGTGTGAAGAAGAAGAACTGAAGCATGACCCCTCTCAAGC 150419
QY 2518 tggctcactcagacacattgagacaacccctggacagccatgcccagagagagagccttga 2577
|||||
Db 150418 TGGCTTAG-----TTGGACAACACCTTGACACAGCCACACAGAGAGAGGCTTCAA 150368
QY 2578 ccggcccccagagctaaagaacacagagagaatacaaatgcttctactcagcgtgacccaa 2637
|||||
Db 150367 CCAGCCCCAGAGCTAAAGCACACAGAGAAAGTCAAAATGCTTCTATTCAAGCATGACCTTAA 150308
QY 2638 ctttctagtgtgcaacgccccacacacctctgcagtaccacacacacacactgctt 2697
|||||
Db 150307 C--TTCTAGTGTGCCAGCGGCCCCACACACTTCTGCAAGTGCCCCACACCATCACCACTGCTT 150250
QY 2698 tctcttcaacagtgatctgtatcttctagtcttcatatcttcttgatgtatat----- 2752
|||||
Db 150249 TCTCTTCCACAGATGATCTGTATTTCTAGTTTCAATTATTTCTTTTATTTGATGACAC 150190
QY 2753 ----- 2752
Db 150189 TATATAAATTTTCTCTCTCTGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 150130
QY 2753 ----- 2752
Db 150129 TTCTTTCTTCAAGACAAGTCTCACTCTGTCAACCCAGGCTGAGTGCAAGGACAGATC 150070
QY 2753 ----- 2752
Db 150069 TTGGCTCCCTGCAACTTCTGCTCCAGGTTCAAGCAATTGTCTGCTCAGCCTCCCGAG 150010
QY 2753 ----- 2752
Db 150009 TAGCTGGGCCACACACCCAGCTAGTTTGTGTACTTTTAATAGAGAGGGGTTCAACAT 149950
QY 2753 ----- 2752
Db 149949 CTTGGCCAGGCTGTCTGAACCTCTGACCTGCTGATCCACCCGCTCGGCTCCCAAG 149890
QY 2753 -----gacactataaanaatttcattga 2777
|||||

/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1.3059
/gene="Stau"
227.1714
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila staufer;
contains four RNA-binding domains; rStau+I6"
/codon_start=1
/product="staufer isoform Stau+I6"
/protein_id="AAF98119.1"
/db_xref="GI:9754866"
/translation="MYKPPVDPHSRMSTVSYGMRGAYPRRYFPYPPVPLLYQVELS
VGGQFNGKGMRPYKHDATARALRTLOEPLPERLEVNGRESEENLNKSEISOVE
EIALKRNLPVNESEFPLTOVARESGPHMKNFVTRVSVGEFVGESEKSKISKNA
RAVLEQLRLPLPAVERVKPRIKKKSQPTCKLQAPDYGQMPISRLAQIOAKKE
KEPEYMLTERGLPRREEFVMOVKVGHHTAGAGTNKKVAKRNAENMLEILGFKVPQ
AQPAPALKSEKTPYKKPGDGRKVTFFEPSGDENGTSNKEDEFKPYLSHQOLPAG
ILPMVEVAQAVGSOGHHTKDFARAAPNPAKATVTAMIARELLYGSTPTAETILKS
NISSGHVPHGPRTRPSEQLYLSRAQGFQVEYKDPKNKNCEVSLINCSOPPLVSH
GIGKDVESCHDMAALNLIKLSLDDQSTEMPRITGNGPVSACGRC"
3043..3048
polyA_signal

BASE COUNT 817 a 778 c 749 g 713 t 2 others
ORIGIN

Query Match 35.8%; Score 1254.8; DB 74; Length 3059;
Best Local Similarity 86.3%; Pred. No. 3.7e-283;
Matches 1429; Conservative 0; Mismatches 207; Indels 20; Gaps 3;

QY 611 cagaagcataccccctactgttagaactaaatgcactgtgcatagaaacttggaaaaaac 670
Db 165 CAGAAAGTATAGCTTCTACCATTTGACCTCAATGCACCTGCGTGAACCTGGAAGAAAC 224
QY 671 caatgataagcctgttgacccttctcgtgagtcagtcacccataactacaacatga 730
Db 225 CAATGTATAAGCCCGTGACCCCACTCTCGATGACATCCACTACGCTATGGCATGC 284
QY 731 gagagtgctatcccccgagtaactttaccattccagttccacttactatc 790
Db 285 GTGAGAGTGCCTATCCCCCAGATCTTTTACCATTTCAGTTCCACCGTTACTCTACC 344
QY 791 aagtgaacttctgtgagagacagcaatttaatgycaaagaaagacaagacagctg 850
Db 345 AAGTTGAGCTCTCCGTGGAGACAGACAGTTTATGGGAAGGAAAGATGAGACCACTG 404
QY 851 cgaacacagatgctgtgcacaaagcgttgagatcctgcagaaatgagccctgcagaga 910
Db 405 TGAACATGATGCCACTGCCCGAGCCCTGAGGACTCTGCACAGTGAAGCCCTTACCAGA 464
QY 911 ggtcgtgagtgaaatggaagaaatccgaagaagaatctcaataaatctgaataagtc 970
Db 465 GGCTAGAGGTTAATGGAAGAAATGGAAGAAACCTTAATTAATCAGAAATTAAGCC 524
QY 971 aagtgttgaagtgcacttaacgaaactgtcctgtgaattcg----- 1015
Db 525 AAGTGTGAGATTGCGCTGAAGCGGAATTGCTGTGAATTGAGTCTTCCCTCTGA 584
QY 1016 ---agtgcccgagagagtggtgccccacacatgaagaacttltgacaaagtctcg 1072
Db 585 CACAGGTGCGCCGGGAGAGTGGCCGCCACACATGAAGAACTTGTGACCAAGGTTTCAG 644
QY 1073 ttggtgagtttgt 1132
Db 645 TTGGGGAGTTTGTAGGGGAAGAGGAAAGCAAGAGATCTCCAAGAGATGCGG 704
QY 1133 ccatagctgttctttaggagctgaagaagttaccgccccctgctgagttgaagagtaa 1192
Db 705 CAAGGGCCGTGCTGAGCAGCTAAGAGGCTGCCACCTTACTGCTGTGAGCGAGTGA 764
QY 1193 agcctagaatcaaaaagaanaaacccatagltcaagccacagacagccagatatg 1252

Db 765 AGCCAGAAATCAGAAGAAAGTACAGCCCACTGCAAGCTACAGACTGCCCCGACTATG 824
QY 1253 gccagggatcaatccgattagccgactgcccagatccagcagggcaaaaaagagagaag 1312
Db 825 GCCAAGGATGATCTTATTAGACAGCTGGCACAGATCCAGCAGGCCAAAAAGAGAGA 884
QY 1313 agccagatgacgctcctcacaagagcgaagcctcccgccgagggaggttgtatgc 1372
Db 885 AGCCGAGTACATGCTCTTACAGACGAGGCTCTTCCACGCTCCAGGAGTGTGATGC 944
QY 1373 agtgaagtttgaaacacacactgcagaaggaagcggcaccacaagaaggttgccaagc 1432
Db 945 AGTGAAGTTCGGCATCACACCGCAGAAGAGCAGGTACCAATAAAGAGTGGCCAAGC 1004
QY 1433 gcaatgcagccgagaacatgtgtgagatccttggttcaaaagtcgagcagccca 1492
Db 1005 GCAATGCTGCTGAGAACATGCTGGAGATCCTGGGTTCAAGTCCCCCAGGCGCAGCTG 1064
QY 1493 ccaaacccgcaatcaagtcagagaggaagacacccataaagaanaaccagggatgaa 1552
Db 1065 CCAAGCCAGCACTCAATCAGAAGAGAGACTCCAGTAAGAACAACAGAGATGGAAGA 1124
QY 1553 aagtaacctttttgaacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1612
Db 1125 AAGTAACATTTTGAACCTAGCCCTGGGATGAATAATGCAACCAAGTAATAAGAGATG 1184
QY 1613 agtcagatgacctatctaatgcatcagcagctgtcctgtgaattcttccatgtgc 1672
Db 1185 AGTTCAAGATGCCCTATCTTAGCCATCAGCAGCTGCCAGCGGAGTTCTCCCATGTGC 1244
QY 1673 ccgagtcgcccaggtctgtagagttagtcagagatcacaccaaagatttaccagg 1732
Db 1245 CAGAGGTTGCCAGGCTGTGCGGGTTAGTCAAGAGACACCAATCAAGATTTCGCCAGG 1304
QY 1733 cagctccgaatcctgccaagggccaaggttaactgcatgtagcccgaggtgtgtatg 1792
Db 1305 CAGCTCCAATCTCTGCCAAGGCCACGGTAAGTCCATGATAGCCGAGATTTGTTAGC 1364
QY 1793 gggcacctgcgccacagccgagacattttaagaataacatctctcagggcagctac 1852
Db 1365 GGGCACCTGCCCCACAGCCGAGACCAATTTAAGAGTAACATCTCTCAGGACATGTAC 1424
QY 1853 cccatgacctctcacgagaccccttgagcaacttggaactatcttccagagtcaggat 1912
Db 1425 CCCATGACCTCGCACAGACCCCTTGAGCAACTGTACTACTCTTCCAGAGCCCAAGGAT 1484
QY 1913 tccaggttgatacaaaagacttccccaaaaaacaacaagaagaattgtatcttataca 1972
Db 1485 TCCAGGTTGAATACAAAGACTTCCCAAGAAACAACAAGAGAGAGTGTATCTCATCA 1544
QY 1973 atgtctctcagccacctgtgatcagccatgtgtatcgtgcaagatgtgtgagtcgcc 2032
Db 1545 ACTGCTTTCACAGCCACCTCTCGTACGCCATGGCATCGGCAAGGATGTGAGTCTGTG 1604
QY 2033 atgatgt 2092
Db 1605 ATGATATGGCTGCACCTGAACATCTTAAGCTGTGTGTGAGTTGGACCAACAGACACAG 1664
QY 2093 agatgccaaagaacaggaagcgaacatgtctgtgtgtgtgtgtgtgtgtgtgtgtgt 2152
Db 1665 AGATGCCAAGGACAGGAACGGAGCACTTTCAGCGTGGCGGAGGTGCTGAACCTATTCTG 1724
QY 2153 gccatgaacattataaaatcccaacatatatactgaataactga-aactgtcttgaaa 2211
Db 1725 GCCATGAACCAATTATAAAA-CCCAACATATATACCGAAAAATACTGAGAAGTGTGAAA 1783
QY 2212 atttgaatttctgatcctccagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2247
Db 1784 ATTTGAATTTCTGATAACTCCAGTGGGCCAAGACA 1819

RESULT 11

Db 285 GTGAGGTG CCTATCCCCCAGATACTTTTACCAATTTCAGTTCACCGTTACTCTACC 344

QY 791 aagtygaacttctctgtggaagacagcaatttaatgycaaaagaaagaaacagacagcgtg 850

Db 345 AAGTTGAGCTCTCCGTGGGAGACAGCAGCTTAAATGGGAAGGAAAGATGAGACCACCTG 404

QY 851 cgaacaacgatgtctgtcccaaacglttagagatccctgcagaatgagccctgcagaga 910

Db 405 TGAACATGATGCCACTGCCAGCCCTGAGGACTCTGCAGAGTGAGAGTGCACCCCTACCAGAGA 464

QY 911 ggcctgaggtgaatggaagagaatccgaagaagaaatctcaataatctgaaataatg 970

Db 465 GGCTAGAGGTTAATGGAAGAAATCAGAGAAGAAACCTTAATAATCAGAAATAAGCC 524

QY 971 aagtgcttgagattgcacttaaacggaacttgccctgtgaatttcgaggtgcccggaga 1030

Db 525 AAGTCTTGAGATTGCCGTGAAGCCGAATTTGCCCTGTGAATTTTGAGGTTGCCCGGGAGA 584

QY 1031 gtgccccacccacatgaagaacttgtgaccaaagtttcggttgaggaggttgaggg 1090

Db 585 GTGCCCCGCCACACATGAAGAACTTTGTGACCAGGGTTTCAGTTGGGGAGTTTGTAGGGG 644

QY 1091 aagtgtaagggaagcaagaagatttcaagaagaaatgycgcacatagctgttcttag 1150

Db 645 AAGGAGAAGGGAAGCAAGAGATCTCCAAGAAAGATGCGCAAGGCGCTGTGAGC 704

QY 1151 agctgaagaagttaacgccccctgctgcagtgaacgagtaaaagcctagaaatcaaaaga 1210

Db 705 AGCTAAGAGAGGCTGCCACTCTACCTGCTGTGAGCGAGTGAAGCCCAAGATCAAGAAGA 764

QY 1211 aaacaaaccatagtcaagccacagacaagcccaagaaatatgcccagggatcaatccga 1270

Db 765 AAGTCAAGCCCACTGCAAGCTACAGACTGCCCCGACTATGCCCAAGGATGAATCCTA 824

QY 1271 ttgaccgactggcccgatccagcagggcaaaaaaaggaagagccagagtacacgctcc 1330

Db 825 TTAGCAGACTGGCACAGATCCAGCAGGCAAAAAAGAGAAGAGCCGGAGTACATGCTCC 884

QY 1331 tcaagagcgaagcctccgcgcgcgcaaggaggtgtgaltgcaggtgaaggttgaacc 1390

Db 885 TTACAGAAAGAGGTCTTCACAGTCGCAGGAGTTTGTGATGCAGGTGAAGGTTGGGCATC 944

QY 1391 aacatgcagaaggaacgycgacccaacaagaaggtgccaagcgaatgcagccgagaaca 1450

Db 945 ACACCGCAGAAGGAGCAGTAACCAATAAGAAAGTGGCCCAAGCCAATGCTGTGAGAGACA 1004

QY 1451 tgcctgagatccttggttccaagtcccgcaagcgcagcccaaccaaccgcactcaagt 1510

Db 1005 TGCTGGAGATCCTGGGGTTCAAGTCCCCCAGGCCGACCTTGCCAAAGCCACTCAAAAT 1064

QY 1511 cagaaggaagaacaccccataaagaacccaagggatggaagaaagtaaccctttttgaac 1570

Db 1065 CAGAAGAGAAGACTCCAGTAAAGAAACCAAGGAGATGGAAGGAAGTAACATTTTGTGAAC 1124

QY 1571 ctggctctggtggaatgaatgggactagtaataagagagatgagttcagagatgcctatc 1630

Db 1125 CTAGCCCTGGGATGAATAAGAACCAAGTAATAAGAGATGAGTTCAGATGCCCTATC 1184

QY 1631 taagtcatcagcagctgcctgctgtaattcttcccatgtgtcccgaggtcgcccgagctg 1690

Db 1185 TTAGCCATCAGCAGCTGCCAGCCGGGATTTCTCCCATGGTGCAGAGAGTTGCCCAAGGCTG 1244

QY 1691 taggaagttagtcaaggaacatcacaccaagaatttaccagggcagctccgaatcctgcca 1750

Db 1245 TCGGGGTTAGTCAAGGACACCATACCAAAAGATTTGCCCAGGGCAGCTCCAATCCTGCCA 1304

QY 1751 aggcacacgtaactgcatgataagccgagagttgtlaltggygcacctcgccacag 1810

Db 1305 AGGCCACGGTAACTGCCATGATAGCCCCGAGAAATTGTTGACGGGGGCACCTCGCCACAG 1364

QY 1811 ccgagaccatttaagaataacatctctcaggccacgtlaccacatgacctctcacga 1870

Db 1365 CCGAGACCATTTTAAAGTAACATCTCTTACAGACATGTACCCCATGACCTCGCACAA 1424

QY 1871 gacctctgagcaactggaactatcttccagaggtccaggatccaggttgaatacaag 1930

Db 1425 GACCTCTGAGCAACTGTACTACCTTTCAGAGCCGAGGATTCAGAGTTGAATACAAAG 1484

QY 1931 acttcccaaaaacacaagaagaacttgtatatctctatcaatltgctcctcgaaccac 1990

Db 1485 ACTTCCCAAGAACACACAGACAGAGTGTGTATCTTCATCACTGCTCTTCACAGCCAC 1544

QY 1991 ctctgatacgcatagtatcggcaagatgtgtgagtcctgcatagatagtgcgtga 2050

Db 1545 CTCTCGTCAGCCATGGCATGGCAAGGATGTGGAGTCTGTCAATGATGGCTGCACCTGA 1604

QY 2051 acatctaaagttgctgtctgaglttgaccacaagtlacagagatgccagaagaa 2110

Db 1605 ACATCTTAAAGCTGCTGTGAGTTGGACCAACAGACACAGAGATGCCAAGAGACAGAA 1664

QY 2111 acggaaccaatgtctgtgtgaggaggtgcgaaccttctgccaatgaaccatataa 2170

Db 1665 ACGGACCAGTTTCAGCGTGGCGGAGGTGCTGAACCTATTTCTGGCCATGAACATTATAA 1724

QY 2171 atcccaacatatatactgaaataactga-aactgctttgaaatlttgaaatttctgatac 2229

Db 1725 A-CCCAACATATATACCGAAATACTGAGAAGTCTTTGAAAAATTTGGAATTTCTGATAA 1783

QY 2230 ctccagtyggccgagaga 2247

Db 1784 CTCAGTGGGCCAAGACA 1801

RESULT 9

AF061942

LOCUS AF061942 2860 bp mRNA ROD 04-MAR-1999

DEFINITION Mus musculus RNA-binding protein staufen mRNA, complete cds.

ACCESSION AF061942

VERSION AF061942.1 GI:4335944

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2860)

AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.

TITLE Mammalian staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum

JOURNAL Mol. Cell Biol. 19 (3), 2220-2230 (1999)

MEDLINE 99147057

REFERENCE 2 (bases 1 to 2860)

AUTHORS Duchaine,T., Luo,M. and DesGroselliers,L.

TITLE Direct Submission

JOURNAL Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900

FEATURES

source

1. 2860

/organism="Mus musculus"

/db_xref="taxon:10090"

326. 1789

/codon_start=1

/product="RNA-binding protein staufen"

/protein_id="AAD17529.1"

/db_xref="GI:4335945"

/translation="MYKPVDPHSMOSTYSYGMRGAYPPRYFPYPVPLLYQVELS

VGGQFNKGKMRPVPKHDAPARALRTLQSEPLPERLEVNGREAEEENLNKSEISQVF

EIALKRNLEPVNEFVARSESPPHMKNFTRVSVGEFVGEGRSKISKNAABAVLEQ

LRLPLPAVERVKPRIKKKSQPTCKTAPDYQGMNPSRLAQIOQAKKEPEYMLL

TERGLPRREEVMQVKGHHTAEVGTNKKVAKRNAENMLEILGFKVPOAPAKPAL

KSEKTPYKHKFDGRKVTFFEPSPDENGTSNKDEFRNPYLSHOQLPAGILPMVPEV

AAVGVSGQHHGDKFTRAAPNPAAKATVTAMIARELLVGTSPTAETILKSNISGHVP

HGPRTRPSEQLYLSRAQGFQVEYKDFPKNNKNECVSLINCSQPLPLVSHGIGKDVES

CHDMAALNLIKLSLDDQSTEMRTGNGPVSACGRC"

BASE COUNT

ORIGIN

747 a 750 c 713 g 650 t

Db	34725	CCCTACGCAATGACAGCTGCGCTGTGGGGAAGAGGGCTGCCACGCCAGCTGTTCTCCCG	34666
QY	2426	ggacaccagcagatccacacccctgagcactccgtgttctgtcttlttcccccgtgt	2485
Db	34665	GGACACCAGCAGATCCACACCCCTGGGCACTCCGTTGTTCCTTTTCCCTGTGT	34606
QY	2486	gaaagaagaaacgycacgaccccttctcaagctggtcactcagaacacattgagacaac	2545
Db	34605	GAAAGAAGAAACGGCAGCACCCTTCTCAAGCTGGCTCAGTCAAGACATTGGAGCAAAC	34546
QY	2546	ccctgacagccatgcccagagagagcctttgaccgccccagagctaaagcaccagaga	2605
Db	34545	CCTGACAGCCCATGCCAGAGAGAGCGCTTGACCGGCCAGAGCTAAAAGCACACAGAGA	34486
QY	2606	aatcaaatgtctcctactcagcgtgaccccaacttctagtgtgccaagcggcccccacac	2665
Db	34485	AAATCAATGCTTCCCTACTCAGCGTGACCCCACTTCTAGTGTGCCACGCGCCACACAC	34426
QY	2666	ctcctgcagtaecccaacaccatcaccaactgtcttcttccaacagtgtatctgtattcta	2725
Db	34425	CTCCTGCAGTACCCACACCATCACCACCTTCTCTCTCTCCAACAGTGTCTGTATTTCTTA	34366
QY	2726	gttcattatttcttcttcttgattgatatgacactatataaaatttcaatttgagaattct	2785
Db	34365	GTTTCATTATTTCTTTTGATGTATGACACTATATAAATTTTTCATTTGAGAAATTTCT	34306
QY	2786	caattgtatctagttaaatagacacagtttgaaaactgtctcagacgtgacttatcaata	2845
Db	34305	CAATTGTATCTAGTTAAATAGCACAGATTGGAAAACTTGTCTGAGACTGACTTTATCAATA	34246
QY	2846	atctaaccgcaaaagatcatatccatgtgtatgtgtagacattttatttcattgact	2905
Db	34245	ATCTAACCGCAAAAGATCATATCCATGTGTATGTGTAGACAAATTTTATTCATTGACT	34186
QY	2906	aaccagagacagtttcagtgatgcaaatgtgtgccctctgtgtcagctgaacagtcct	2965
Db	34185	AAACCAGGACAGTTTCAGTGATGCAAAATGTGTGCCCTCTGGTTCAGCTGAACAGTCCCT	34126
QY	2966	ggacttccaacaaacctgaataagctctcccacagttgtatataattggaacatttagaat	3025
Db	34125	GGACTTTCAAAAACCTTGAATAGTCTCCACAGTTGTATATAATTGGACAATTTAGGAAT	34066
QY	3026	tttaaaccttagatgatacatttgggtccatttltatttcatttltatttltgtlaatgca	3085
Db	34065	TTTAACTTTAGATGATCATTTTGGTTCATTTTATTTTCAATTTTATTTTGTTAATGCA	34006
QY	3086	aacagagacttaaatgaacttgcatactctgttttaagaattatlaaaaaacattgtgtatc	3145
Db	34005	AAACAGACTTAATGAACCTTTGATCTGTGTTTAAAGATTATTAATAAACATTGTGTATC	33946
QY	3146	tatacatatgctcttgagagacttagctttcaactacactacagagatatgatctccatgta	3205
Db	33945	TATACATATGGCTCTTGAGGACTTAGCTTTCACACTACAGATATGATCTCCATGTA	33886
QY	3206	gtccatatataacctcagagtgatatttccagagtgctcgatactgtttaattacatctcca	3265
Db	33885	GTCCATATAAACCTGCAGAGTGATTTTCCAGAGTGCTCGATACTGTTAATTACATCTCCA	33826
QY	3266	ttagggctgaaaaaatgacctacgtttctgtatatacagctgtgtgtcttltgatgtgtg	3325
Db	33825	TTAGGGCTGAAAAAGATGACCTACGTTTCTGTATACAGCTGTGTGCTTTTGATGTGTG	33766
QY	3326	ttactgtacacagaagtgtgtcactgaggtctctgctgtgtgtcctgtatgaaaaacctg	3385
Db	33765	TTACTGTACACAGAGTGTGTGACACTGAGGCTCTGCGTGTGTCGATGGAAGCCTGG	33706
QY	3386	tagccctgaggttaagtaactgtcttccattcatgtgtttagcgtggaatttttctcccat	3445
Db	33705	TAGCCCTGCGAGTTAAGTACTGCTTCCATTCATTGTATTACGCTGGAAATTTTCTCCCAT	33646
QY	3446	ggaatgtlaagtaaaacttaagtgttctgtatcaataaatgtataactaaa	3496

Db	33645	GGATGTAGTAATAAACTTAAGTGTGTCATCATATAAATGGTAATACTAA	33595
RESULT	8		
AF290989			
LOCUS	AF290989	3041 bp	mRNA
DEFINITION	Rattus norvegicus staufer isoform Stau-16 (Stau) mRNA, complete cds, alternatively spliced.		
ACCESSION	AF290989		
VERSION	AF290989.1	GI:9755162	
KEYWORDS			
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	1 (bases 1 to 3041)		
AUTHORS	Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desgroselliers, L., Kuhl, D., Richter, D. and Kindler, S.		
TITLE	Two rat brain staufer isoforms differentially bind RNA		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3041)		
AUTHORS	Monshausen, M. and Kindler, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-JUL-2000) Institute for Cell Biochemistry and Clinical Neurobiology, University of Hamburg, University Hospital Eppendorf, Sudefeldstr. 24, Hamburg D-22529, Germany		
FEATURES			
source	location/Qualifiers		
	1..3041		
	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	/tissue_type="brain"		
	/dev_stage="adult"		
	1..3041		
	/gene="Stau"		
	227..1696		
	/gene="Stau"		
	/note="RNA-binding protein; similar to Drosophila staufer; contains four RNA-binding domains; rStau-16"		
	/codon_start=1		
	/product="staufer isoform Stau-16"		
	/protein_id="AA98177.1"		
	/db_xref="GI:9755163"		
	/translation="MYKPVDPHSRMQSTYSYGMRGAGYPPRYFYPDPVPLLYQVELS		
	VGGQJFNGKMRPVPVKHDATAALRLTLQSEPLERLEVNGRESEENLNKSEISOVE		
	EIALKRLNLPVNEFVARESGPPHMKNFVTRVSGEFVGESEKSKIKNAARAVLEQ		
	LRLPLPAVERVKPRIKKSOPTCKLQTPAPYGGGMPISRLAQIQAKKEPEYIM		
	LTTERGLPRRREFVMQYVGHHTAEGAGTNKVAKNAAEMLEILGFKVPOQAPAK		
	ALKSEKTPVKKPGDGRKVTFFEEPSGDENGTSNKEDEFMPYLSHQOLPAGILPMVP		
	EVAQAVGVSGHHTKDFARAAPNPAPKATVTAMIARELLYGTSPTAETILKSNISSGH		
	VPHGPRTRPSEQLYVLSRAQGFQVEYKDFPKNKNKNECVSLINCSSQDPLVSHGIGKDV		
	ESCHDMAALNLIKLLSELDDQSTEMPRTGNGPVSAACGRC"		
	3025..3030		
	polyA_signal		
	/gene="Stau"		
BASE COUNT	814 a	771 c	747 g
ORIGIN		707 t	2 others
Query Match	36.6%;	Score 1282.8;	DB 74; Length 3041;
Best Local Similarity	87.2%;	Pred. No. 1e-289;	
Matches 1429;	Conservative	0;	Mismatches 207; Indels 2; Gaps 2;
QY	611	cagaagcataaccctactgtagaactaaatgcactgtgcatgaaacttgaaaaaac	670
Db	165	CAGAAAGTATAGCTTCTACCATTTGAGCTCAATGACACTGTGCGTGAACACTGGAAGAAAC	224
QY	671	caatgtataagcctgttgacccttactctcggatgacgtccacactataactacaatga	730
Db	225	CAATGTATAAGCCCGTGGACCCACACTCTCGGATGCAGTCCACTACAGCTATGGCATGC	284
QY	731	gaggagtgcttaccctcccgaggtacttaccattccagttccaccttacttacc	790

Db 2037 CGGCCCCACCACCTCGCAGTACCCACACCATCACCTGCTTCTCTCCACACAGTGA 2096

QY 2714 tctgtatctttagtttcatatttcttcttgattgatatgacactatataaaatttcat 2773

Db 2097 TCTGTATCTTAGTTTCATATTTCTTTTGATGTGATATGACACTATATAAATTTTCAT 2156

QY 2774 ttgagaatttctcaattgtatctaglttaaatagacagtttggaaactgtctgagactg 2833

Db 2157 TTGAGAAATTTCTCAATTTGATCTAGTTAAATAGACAGTTTGAAACCTGTCTGAGACTG 2216

QY 2834 acctataaataaactaaccgacaaagatcatatcatgtgtatgtgttagacattt 2893

Db 2217 ACTTTATCAATATCTAACCGCAAAAGATCATATCCATGTGTATGTGTAGACATTTT 2276

QY 2894 attcattgactaaccagagacagtttcagtgatgcaaatgtgtgcccctggtcagc 2953

Db 2277 ATTTCATTGACTAACCCAGACAGTTTCAGTGATGCAAAATTTGTGCCCTGTGTTGAGC 2336

QY 2954 tgaacagtcctgagacttccaacaaacctgtaataagtcctccacagttgtataattgga 3013

Db 2337 TGAACAGTCCTGGACTTTCAAAAACCTGGAATAGTCTCCACAGTGTATATAATTGGA 2396

QY 3014 caattaggaattttaaacttagatgatcatctgttccatttatttatttattt 3073

Db 2397 CAATTTAGGAATTTTAACTTTAGATGATCATTTGTTCCATTTTATTTTATTTTATTT 2456

QY 3074 ttgttaatgcaacagagacttaaaatgaacttgcattctgtttaaagattataaaa 3133

Db 2457 TTTGTTAATGCAAAACAGAGACTTAAATGAACCTTGTGATCTCTGTTTAAAGATTATTAAAA 2516

QY 3134 acattgttatctatatacatatgctcttgagacttagcttccactacactacagatat 3193

Db 2517 ACATGTGTATCTATATACATATGGCTCTTGAGGACTTAGCTTTCACACTACACAGATAT 2576

QY 3194 gatctccatgttagtccatataaaacctgcagagtgatgttccagagtgctcgatactglt 3253

Db 2577 GATCTCCATGTAGTCCATATTAACCTGCAGAGTGATTTTCCAGAGTGCTGCATACTGTTA 2636

QY 3254 attacatctccattagggctgtaaaagaatgacctacgttctgtatatacagctgtgtgct 3313

Db 2637 ATTACATCTCCATTAGGGCTGAAAGAATGACCTAGCTTTCTGTATACAGCTGTGTGCT 2696

QY 3314 ttgatgtgtgttactgttaccagagaagtgtgtgcactgagagctctgctgtgtgtcgtga 3373

Db 2697 TTTGATGTGTGTACTGTACACAGAAGTGTGTGCACTGAGGCTCTGCGTGTGCTCCGTA 2756

QY 3374 tggaaaacctgtagccctgcgagtgtaagtagtcttccattcatctgtttaagctggaat 3433

Db 2757 TGGAAAGCCTGTAGCCCTGCGAGTTAAGTACTGCTTCCATTCATTTGTTTACGCTGGAAT 2816

QY 3434 tttctcccatggaatgttaagtaaaacttaagtgttgcatacatcaataaagttaatact 3493

Db 2817 TTTTCTCCCATGGAATGTAAAGTAAAACTTAAGTGTGTCATCAATAAATGTAATACT 2876

QY 3494 aaa 3496

Db 2877 AAA 2879

RESULT 7

AL13174/c 113836 bp DNA PRI 14-JUN-2000

LOCUS

DEFINITION Human DNA sequence from clone RP3-470L14 on chromosome 20, complete sequence.

ACCESSION AL13174

VERSION AL13174.15 GI:8573761

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 113836)

AUTHORS Ramsay,H.

TITLE Direct Submission

JOURNAL Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT requests: clonerquest@sanger.ac.uk

On Jun 20, 2000 this sequence version replaced gi:8388434.

During sequence assembly data is compared from overlapping clones. Where differences are found there are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20

RP3-470L14 is from the library RPI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-470L14 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP3-470L14 is at 113836 in this sequence. The true right end of clone RPI-155G6 is at 100 in this sequence.

FEATURES

source

1. .113836

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/clone="RP3-470L14"

/clone_1kb="RPI-3"

/clone_1lb="RPI-3"

/note="weak data"

95583. .95627

/note="Single clone region. pUC subcloned to give 12x coverage. Assembly confirmed by restriction digest data."

BASE COUNT 33974 a 25115 c 24233 g 30514 t

ORIGIN

Query Match 38.7%; Score 1357.4; DB 34; Length 113836;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2127 gtgtggaggtgctgaaccttcttgccatggaacattataaaatcccaatatatac 2186

Db 34965 GTGTGGAGGTGCTGAACCTTTCTGGCCATGAACCATTTAAATCCCAATATATATAC 34906

QY 2187 tgaataactgaactgcttgaanaatttggaattctgatacctccagtggtgcccagag 2246

Db 34905 TGAATAACTGAACACTGCTTGAATAATTGGAATTTCTGATACCTCCAGTGGCCGAGAG 34846

QY 2247 acacggtggttaagaagtgtggcagcagcaggggaagacaacagaacacaaggagcgg 2306

Db 34845 ACACGGTGGGTTAAAGATGTGGCCAGCAGCAGGGAAGACAACAGAAACAGAGAGCGG 34786

QY 2307 ctgtggcc-gcctgagactgtgtgggtgtgtgtatgtgccaactcggtgtaacctggtg 2365

Db 34785 CTGTGGCCGGGCTGGACTGTGCTGGGTTGTGTGATGGCCACTCGGTGACTGGCGGT 34726

QY 2366 ccctaagcaatagcagctgtgtgtgggaaggagctgcaccagccagctgttctcccg 2425

BASE COUNT 891 a 693 c 699 g 783 t
ORIGIN

Query Match 79.8%; Score 2799.4; DB 68; Length 3066;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2870; Conservative 0; Mismatches 6; Indels 7; Gaps 6;

QY 617 gcataaccctctgtagaactaaatgcaactgtgcataaacttggaaaaaaaccatgt 676
Db 1 GCATAACCCCTACTGTAGAACTAAATGCACTGTGCATGAACCTTGAAAAAACCAATGT 60
QY 677 ataagcctgttgaacccttactctcgtgagtcagtlccacctataactacaacatgagagag 736
Db 61 ATAAGCCTGTGACCCTTACTCTCGGATGAGTCACCTATACTACAACATGAGAGAG 120
QY 737 gtgtctatcccccgaggtaacttttaaccaattccagtlccaccttacttatacaagtgt 796
Db 121 GTGCTTATCCCCGAGTACTTTTACCATTCCAGTTCACCTTTACTTTATCAAGTGG 180
QY 797 aacttctgtgagagacagcaatttaatgtgcaagaagaacaagacagagctgcgaac 856
Db 181 AACTTCTGTGGAGGACAGCAATTTAATGGCAAAGGAAGACAGACAGGCTGCCGAAC 240
QY 857 acgatgtctgtcccaaaagcgtltgagatcctctgcagaatgagccctgtccagagagctgtg 916
Db 241 ACGATGCTGCTGCCAAAGCGTGTGAGGATCTCGAGAATGAGCCCTGCCAGAGGCTGG 300
QY 917 aggtgaatggaagaagaatccgaaagaaaaaatctcaataatctgaaaaataagtcgaagt 976
Db 301 AGGTGAATGGAAGAGAAATCCGAAGAAGAAAAATCTCAATAAATCTGAATTAAGTCAAGTGT 360
QY 977 ttgagatgtcaactaaacggaactgtgcctgtgaaatlccgagtgccccggagagtgcc 1036
Db 361 TTGAGATTGCATTAAACGGAACTTGCTGTGAATTTGAGGTGGCCCGGAGAGTGGCC 420
QY 1037 cacccacatgaagaacttltgaccaaggtttcgtgtggaggttltgtggggaagtg 1096
Db 421 CACCCACATGAAGAACTTTGTGACCAAGGTTTCGGTTGGGAGTTGTGGGGAAGGTG 480
QY 1097 aaggaaaaagcaagaagatttcaagaanaaatgccccatagctgttcttgagagctga 1156
Db 481 AAGGAAAAAGCAAGAAATTTCAAGAAAAATGCCGCCATAGCTTTCTTGAGGAGCTGA 540
QY 1157 agaagttaccgccccctgctgcagtttgaacgagtaaagcctagaatcaaaaaagaacaa 1216
Db 541 AGAAGTTACCGCCCCCTGCCTGCAGTTGAACGAGTAAGCCTAGATCAAAAAAGAAACAA 600
QY 1217 aaccatagtcaagccacagacaagaagccccagaatatgccaagggatcaatccgattagcc 1276
Db 601 AACCATAGTCAAGCCACAGACAAGCCCCCAAGATATGGCCAGGGGATCAATCCGATTAGCC 660
QY 1277 gactggccagatcccaagcaggcaaaaaagagaagagccagagtaacaagctcctcacag 1336
Db 661 GACTGGCCAGATCCAGCAGGCAAAAAAGAGAAAGAGCCAGAGTACACGCTCCTCACAG 720
QY 1337 agcgaagcctcccgccgcccagggaggtltgtgatgcaggtgaaggttggaaaccacactg 1396
Db 721 AGCGAGGCTCCCGCGCCGAGGAGTTTGATGTCAGGTGAAGTTGGAACCAACACTG 780
QY 1397 cagaaggaacgagccaacaagaaggtgccaagcgcaatgcagccgagaaacatgtctg 1456
Db 781 CAGAAGGAACGGGCAACAAGAAGGTGGCCAAAGCGCAATGCAGCCGAGAACATGCTGG 840
QY 1457 agatccttgttttcaaaagtcocccgacgagcccaaccaaccgcaactcaagtcaaggt 1516
Db 841 AGATCCTTGGTTTCAAGTCCCGCAGCG--CAGCCACCAAAACCCGCACCTCAAGTCAGAGG 898
QY 1517 agaagacacccataaagaaaaccaaaggtggaagaaagaagtaacc-tttttgaacctgac 1575
Db 899 AGAAGACACCCATTAAGAAACCAAGGGGATGGAAGAAAGTAAACCTTTTGTGAACCTGGC 958
QY 1576 tctgggat-gaaaaatggaactagtaataaagaagatgagtlccagatgacctatactaa 1634

Db 959 TCTGGGATGGAATAAGGACTAGTAATAAGAGGATGAGTTTCAGGATGCCCTATCTAAG 1018
QY 1635 tcatcagcagctgtcctgtgnaattcttcccatgtgtcccgagtcgcccagctgtag 1694
Db 1019 TCATCAGCAGCTGCTGCTGGAATCTTCCCATGTGTGCCGAGGTGCCCCAGGCTGTAG 1078
QY 1695 agttagtcaagacatcacaccaagaagatttaaccagggcagctccgaatccctgcgaagc 1754
Db 1079 AGTTAGTCAAGGACATCACACCAAGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGC 1138
QY 1755 caggtlaactgccatgatagcccagaggttgttaltgtggygcacctcgcacagccga 1814
Db 1139 CACGGTAAGTGCCTGATAGCCCCGAGAGTGTGTATGGGGGCACCTCGCCACAGCCGA 1198
QY 1815 gaccattttaaagaataacatctctcaggccaagtcaccccatggaacctctcagagacc 1874
Db 1199 GACCATTTTAAGAAATTAACATCTCTTCAGGGCCACGTAACCCCATGGAACCTCTCACGAGACC 1258
QY 1875 ctctgacaaactgactalcctccagagtcaggatccaggttgaatataaagactt 1934
Db 1259 CTCTGAGCAACTGCACTATCTTTCCAGAGTCCAGGGATTCCAGGTGAATACAAAGACTT 1318
QY 1935 ccccaaaaacaagaagaacttltgatccttatcaaatgtctcctcagccactct 1994
Db 1319 CCCCAAAACAACAAGAACGAATTTGTATCTTATCAATTTGCTCTCAGCCACCTCT 1378
QY 1995 gatcagccatgtatcggcaagatgtgtgagtcctcgcatgatgtgctgcgtgaacat 2054
Db 1379 GATCAGCCATGGTATCGGCAAGATGTGAGTCCCTGCCATGATATGGCTGCGTGAACAT 1438
QY 2055 cttaaagttgctgtctgtgaglttggaccaacaagltacagagatgcgaagaacaggaacg 2114
Db 1439 CTTAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGACAGGAACGG 1498
QY 2115 accaatgtctgtgtgtggaggtgtcgtgaaccttltctygccatgaaacctataaatcc 2174
Db 1499 ACCAATGCTGTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAACCATTTATAAATCC 1558
QY 2175 caacatatatactaaaaatactgaaactgtcttggaaaatttggaaattctgatacctcca 2234
Db 1559 CAACATATATACTGAATAATACTGAAACTGCTTTGAAAAATTGGAAATTTCGTATACCTCCA 1618
QY 2235 gtggccgagagacacggttggtaaagatgttggcagcagcaggaagaacaacagaac 2294
Db 1619 GTGGCCGAGAGACACGGTGGGTAAAGGATGTGGCAGCAGCAGGGAAGACAACAGAAAC 1678
QY 2295 acaaggaagcggtgtggcc-ggctygactgtgtggggttgtgtgtagtgccactcg 2353
Db 1679 ACAAGGAGGCGGCTGTGGCCGGGCTGACATGTGGGGGTTGTGTGATGGCCACTCGG 1738
QY 2354 tgacctygcggtccctactacgcaatagcagctgtcctgttgggaagaagggctgccagccag 2413
Db 1739 TGACCTGGCGGTCCCTACGCAATAGCAAGCTGCTGTGGGGAAGAGGGGCTGCCACGCCAG 1798
QY 2414 ctggttctcccggaaccaagcagatcccaacccctgggcacctccgtgttgtgtctt 2473
Db 1799 CTGTTCTCCCGGACACCAAGATCCACACCCTGGGCACCTCCGTGTGTTT 1858
QY 2474 ttccacctgtgtgaagaagaacacgycagaccccccttctcaagctgtgtcactcagacac 2533
Db 1859 TTTCCCTGTGTGAAGAAGAAACGGCACGACCCCTTCTCAAGCTGGCTCACTCAGACAC 1918
QY 2534 attggacaacaccttgaacagccatgtccagagagagcccttgaacggtcccagagctaa 2593
Db 1919 ATTGGACAAA-CCTTGACAGCCATGCGCAGAGAGAGG-CTTTGACCGGCCCCAGAGCTAA 1976
QY 2594 aagcaccagagaaaatacaaatgtctcctactcaagcgtgaaccaactttctagtgcca 2653
Db 1977 AAGCACAGAGAAATCAATGCTTCTACTCAGCGTGACCCAACTTTCTAGTGTGCCA 2036
QY 2654 cggccccaccacctcctgtcagtlaccacacatcaccactgtcttcttccacagtg 2713

QY 2052 catcttaaagttgtctgtcgtgagttgagccaacaagtacagagatgccaagaacagaaa 2111
Db 1730 CATCTTAAAGTTGCTGTGTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAGAAA 1789
QY 2112 cggaccaatgtctgtgtgtggtgaggtgtgtgaacctttctgcatgaacctataaa 2171
Db 1790 CGGACCAATGCTGTGTGTGGAGGTGCTGAACCTTTCTGGCCATGAACATTATAAA 1849
QY 2172 tcccaacatatatactgtaaaaatactgaaactgctttgaaaatttgaaatttctgatact 2231
Db 1850 TCCCAACATATATACTGAAATACTGAAACTGCTTTGAAAATTGGAATTTCTGATACCT 1909
QY 2232 ccagtggtgagcgaagacacggttggtaaaagtgtggcagcagcaggaagaacaacaga 2291
Db 1910 CCAGTGGGCCGAGAGACACGGTGGTAAGGATGGGGCAGCAGCAGGGAAGACAACAGA 1969
QY 2292 aacacaagagcgcgctgtggcc-ggctggaactgtgtgggttgtgtgtagtgccact 2350
Db 1970 AACACAAGGAGCGCGGTGTGGCCGGCTGGACTGTCTGGGGTTGTGTGATGGCCACT 2029
QY 2351 cgttgacctggtgtccctacgcaatagcagctgctgtggggaagaagcgctgccagc 2410
Db 2030 CGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCTGTGGGAAGAAGGGTGCCACAGC 2089
QY 2411 cagctgttctcccgagacaccagcagatccacacctgtggcaacctccgtgttgtctt 2470
Db 2090 CAGCTGGTCTCTCCCGGAGACCAAGCATCCACACCTGGGGCACCCTCCGTGTTGCTT 2149
QY 2471 tttttccctgtgtgtgaaagaagaacggaacaccttctcaagctgtgtcactcaga 2530
Db 2150 TTTTTCCTCTGTGTGAAGAAGAAGAACGGCAGCAGCCCTTCTCAAGCTGCTTACTCAGA 2209
QY 2531 cacattggagacaacacctgtgacagccatgtccagagagagcccttgaccggcccccagagc 2590
Db 2210 CACATTGGGACAAACCTGTGACAGCCCATGCCAGAGAGGCGCTTTGACCGGCCCCAGAGC 2269
QY 2591 taaaagcaccagagaaatcaaatgtcttctactacagcgtgacccaacttcttagtgtg 2650
Db 2270 TAAAGCACCCAGAGAAAATCAAAATGCTTCTTACTAGCGTGACCCAACTTTCTAGTGTG 2329
QY 2651 ccacggccccaccacctctgtcagtlaccacacacatcaccactgttctcttccaacag 2710
Db 2330 CCACGGCCCCACACCTCTGTGAGTACCCACACCATCACCACTGCTTCTTCCAACAG 2389
QY 2711 tgatctgtatctttagtttcaatatttcttcttgatgatagcactatataaaattt 2770
Db 2390 TGATCTGTATTCTTAGTTCAATTATTTCTTTTGATGATATGACACTATATAAAATTTT 2449
QY 2771 catttgagaattctcaattgtatctagttaaatagcacagtttggaaactgtctgaga 2830
Db 2450 CATTTGAGAAATTTCTCAATGTATCTAGTTAAATAGCACAGTTTGGAAACTGTCTGAGA 2509
QY 2831 ctgacttatcaataatctaaccgacaagaatcatatccatgtgtatgtgttagacatt 2890
Db 2510 CTGACTTTATCAATAATCTAACCCGACAAAGATCATATCCATGTATGTGTTAGACATT 2569
QY 2891 ttatattcattgactaacccagagacagtttcagtgatgcaaatgtgtgccctctgttc 2950
Db 2570 TTTATTTTCATGTAGCTAACCCAGACAGTTCAGTGTGCAAAATGTGTGCCCTCTGGTTC 2629
QY 2951 agctgaaaacagtcctgtgactttcaaaaaccttgaaataagtcctccacagttgtataaatt 3010
Db 2630 AGCTGAAACAGTCTCTGAGCTTTCAAAAACCTTGAAATAGTCTCCACAGTGTATTAATT 2689
QY 3011 ggaacaatttaggaattttaaactttagatgatcatctgttccattttatattt 3070
Db 2690 GGACAATTTAGGAATTTTAAACTTTAGATGATCATTTGGTTCATTTTATTTCAATTTT 2749
QY 3071 atttttgttaatgcaaaacggacttaaatgaacttgtatctctgtttttaaagattataa 3130
Db 2750 ACTTTTGTAAATGCAAAACGAGACTTAAATGAACTTTGATCTCTGTGTTTAAAGATTATTAA 2809

QY 3131 aaaacattgttatctatacatatgctcttgaggacttagctttcaactaacacagga 3190
Db 2810 AAAACATTGTGTATCTATATACATATGGCTCTTGAGGACTTAGCTTTCACCTACACAGGA 2869
QY 3191 tatgaltccatgtagtlccataataacctgcagagtgatttccagagtgctcgaactg 3250
Db 2870 TATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGATTTTCCAGAGTGTCTGATACTG 2929
QY 3251 ttaatacatctccatlagggctgaaaagaatgactaagttctgtatatacagctgtt 3310
Db 2930 TTTAATTACATCTCCATTAGGGCTGAAAAGAATGACCTACGTTTCTGTATACGCTGTGTT 2989
QY 3311 gctttgattgtgtgttactgtacacagaagtggtgcactgaggtctgcgtgtgtcc 3370
Db 2990 GCTTTGATGTGTGTACTGTACACAGAGAGTGTGTCACTGAGGCTCTGCGTGTGTC 3049
QY 3371 gtatgaaaacctggtagccctgcgagtttaagtactgtctccattcatgtttaacgtg 3430
Db 3050 GTATGGAAGCCTGCTAGCCCTCGAGTTAAGTACTGCTTCATTCATTGTTACGCTGG 3109
QY 3431 aatttttctcccatggaatgtaaagtaaaactaagtgttgtcatcataaatgttaat 3490
Db 3110 AATTTTCTCCCATGGAATGTAAGTAAAGACTTAAGTGTGTTGTATCATATAATGTGAAT 3169
QY 3491 actaaaaaaaaaaaaa 3506
Db 3170 ACTAAAAAATAAAAAA 3185

RESULT 6
HSA132258
LOCUS HSA132258 3066 bp mRNA PRI 14-MAY-1999
DEFINITION Homo sapiens mRNA for stauflen protein, partial.
ACCESSION AJ132258
VERSION AJ132258.1 GI:4572587
KEYWORDS stauflen gene; stauflen protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3066)
AUTHORS Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
TITLE Interaction of influenza virus NS1 protein and the human homologue
of Stauflen in vivo and in vitro
JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE 99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS Ortin,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,
Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN
FEATURES
source location/Qualifiers
1. 3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
36. 1526
/gene="stauflen"
36. 1526
/gene="stauflen"
/function="dsRNA binding protein"
/codon_start=1
/product="stauflen protein"
/protein_id="CAB40082.1"
/db_xref="GI:4572588"
/translation="MKLGKKPMYKPYDSRMOSTYNNYNRGAYPFRYFPVPPL
LYQVELSVGGOQFNGKGTROAKHDAAKALRILONEPLPERLEVNRESEENLNK
SEISOVEFIALKRNLPVNFVARRESGPPHMKNFVTKSVGEFVGEGRSKISKNA
AIAVLELEKLRLPPLPAVERVKPRIKKTKPIVKPOTSPEYGOINPISRLAOIOAKK
EKEPEYTLTERGLPRRREFVMQYKGNHTAEGTNNKYAKRNAENMLEILFRVP
QRSHQRTQVRGEDTHKETRGWKSKNLFNLALGMENGTSNKDEFRMPYLSHQOLPA
GILPMVPEVAQAVGSQGHTKDFTRAAPNPAKATVTAMARELLYGTSPTAETILK
NNISSGHVPHGPLTRPSEQLDLYSRVQGFQVEYKDFPKNNKNEFVLSINCSSQPLIS
HGIGKDVESCHDMAALNILKLSELDOQSTEMPRTGNGPMSVCGRC"

ACCESSION AK025519
VERSION AK025519.1 GI:10438059
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP clone:HEP02379.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hiki,j,i,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE MEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 3194)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT MEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES
source 1..3194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP02379"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
misc_feature 1..3194
/note="highly similar to AF061938 Homo sapiens staufer protein (STAU) mRNA"
BASE COUNT 894 a 787 c 732 g 781 t
ORIGIN
Query Match 81.3%; Score 2848.8; DB 34; Length 3194;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2886; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Db 592 GCTGAGGTGAATGGAAGAATCCGAAGAAGAAATCTCAATAAATCTGAATAAGTCA 651
QY 972 agtgtttagatgtgaacttaaacggaacttgctgtgaatttcgaggttgcccgagag 1031
Db 652 AGTGTGTGAGATGTCATTAACGGAACTTGCCCTGTGAATTTGAGGTGGCCGGGAGAG 711
QY 1032 tggccacccccacatgaagaactttgtgaccaagtttcggttgaggagtttgtgggga 1091
Db 712 TGGCCACCCACATGAAGAACTTTGTGACCAAGGTTTCGGTTGGGAGTTGTGGGGGA 771
QY 1092 agtgtgaaggaagaagaagaatttcaagaaaaaatgcccataagctgttcttgagga 1151
Db 772 AGGTGAAGGGAAGCAAGAAGATTTCAGAAAAATGCCGCCAATAGCTGTTCTTGAGGA 831
QY 1152 gctgaagaagtaccgccccctgctgcagttgaacgagtaaacctagaatcaaaagaa 1211
Db 832 GCTGAAGAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAAGCCTAGAATCAAAAAAGAA 891
QY 1212 aacaaaccatagtcaagccacagacaagcccaaatatgtgccaggggatcatccgat 1271
Db 892 AACAAAA-CCATAGTCAAGCCACAGACAGACCCCAAGATATGGCCAGGGATCAATCCGAT 950
QY 1272 tagccgactgcccagatccagcagggcaaaaaaggaagagagccagagttacagctcct 1331
Db 951 TAGCCGACTGGCCAGATCCAGCAGGCAAAAAAGGAAGAGAGCCAGAGTACACGCTCCT 1010
QY 1332 cacagaagcagcctcccgccgcccagcagggagttgtatgcaggtgaagttgaaacca 1391
Db 1011 CACAGAGCGAGCCCTCCCGCCGCGAGGAGTTTGTGATGCAGGTGAAGTTGAAACCA 1070
QY 1392 cactgcaagaagcagggcaccacaagaaggttggccaagcgcactcagccgagaaacat 1451
Db 1071 CACTGCAGAAGAAGCGGCACCAACAAGAAGTTGGCCAAGCGCAATGCAGCCGAGAACAT 1130
QY 1452 gctgagatccttgttttcaaagtcccgagcggcagcccaaacccgactcaagtc 1511
Db 1131 GCTGGAGATCCTTGTTTCAAGTCCCGCAGCGGCACAGCCCAACCCGCACCTCAAGTC 1190
QY 1512 agaggaagaagacaccataaagaaccagggatgtgaagaaaaagtaacctttttgaacc 1571
Db 1191 AGAGGAGAAGACACCCATTAAGAAACCA-GGGATGGAAGAAAGTAACCTTTTGAACC 1249
QY 1572 tggctctgggagtaaaaaatggaactagtaataaagagatgattcaggatgcctatct 1631
Db 1250 TGGCTCTGGGATGAATAATGGGACTAGTAATAAGAGATGAGTTTCAGGATGCCCTATCT 1309
QY 1632 aagtcacagcagctgcctgtcgtgaattctcccatggtgcccaggtgcgccagctgt 1691
Db 1310 AAGTCATCAGCAGCTGCCTGCTGGAATTTCCCATGTGTGCCGAGGTGCCAGGCTGT 1369
QY 1692 aggagttagtcgaagacatcacaccaagaatttaccagggcagctccgaatccttgcca 1751
Db 1370 AGGAGTTAGTCAAGGACATCACACCAAGAATTTTACCAGGGCAGCTCCGAATCCTGCCAA 1429
QY 1752 ggcacggttaactgcatgatagcccgaagttgtgtatggyggcaacctgcgccacagc 1811
Db 1430 GGCCACGGTAACCTGCATGATAGCCCGAGAGTTGTGTATGGGGGACCTCGCCACAGC 1489
QY 1812 cgagaccattttaagaataacatctctcagggccaagtlaccccatygaacctctcagag 1871
Db 1490 CGAGACCATTTTAAAGATTAACACCTCTTTCAGGCCACAGTACCCTCATGGAACCTCTCAGCAG 1549
QY 1872 accctctgagcaactygaactatcttccagagttccaggatccaggttgaatacaaga 1931
Db 1550 ACCCTCTGAGCAACTGACCTATCTTTCAGAGTCCAGGAGATTCCAGGTTGAATACAAAGA 1609
QY 1932 ctcccccaaaaacaagaagaagaaatttgtatctctatcaatgtctcctcagccacc 1991
Db 1610 CCTCCCAAAAAACAACAAGAAGAAATTGTATCTCTTATCAATGCTCTCAGCCACC 1669
QY 1992 tctgatacgcatagtatcggcaagatgtgaggttcctgacatgatatggtcgtcgtgaa 2051
Db 1670 TCTGATCAGCCATGGTATTCGGCAAGGATGTGAGTCTGTCATGATATGGCTGCGCTGAA 1729

QY 1451 tgcctgagatccttggttccaagtcccgagcgcgcaagcccacaaaccgcactcaagt 1510
|||||
Db 1087 TGCTGGAGATCCTTGTTCAAGTCCCGCAGCGGCAGCCCAACCCGCACTCAAGT 1146
QY 1511 cagaagagaagacaccccataaagaacacgaggtgaaagaaagtlaaccttttgaac 1570
|||||
Db 1147 CAGAGGAGAGACACCCCATTAAGAAACCGAGGGGATGGAAGAAAGTAACCTTTTGAAC 1206
QY 1571 ctgctctggtgatgaaaatgggactagtaataaagagatgattcagagtccctatc 1630
|||||
Db 1207 CTGGCTCTGGGATGAAAAATGGGACTAGTAATAAGAGGATGAGTTCAGGATGCCCTTATC 1266
QY 1631 taagtcatacgcagctgcctgctggaattcttcccatggtgcccgaggtgcgccagctg 1690
|||||
Db 1267 TAAGTCATCAGCAGCTGCCCTGCTGGAATCTTCCCATGGTGGCCGAGGTGCGCCAGGCTG 1326
QY 1691 taggagttagtcgaagacatcacaccaaagaatttaccagggcagctccgaatccctgcca 1750
|||||
Db 1327 TAGGAGTTAGTCAAGGACATCACACCAAGATTTTACCAGGGCAGCTCCGAATCTCGCCA 1386
QY 1751 aggcacacggtlaactgccaatagtagccgagaagttgtgtatggtgggcaacctgcgccacag 1810
|||||
Db 1387 AGGCCACGGTAACTGCCATGATAGCCCCGAGAGTGTGTATGGGGCACCCTGCCCCACAG 1446
QY 1811 ccgagaccattttaagaataacalcctcgaagcagtlaccccatgacctctcaca 1870
|||||
Db 1447 CCGAGACCATTTTAAAGATTAACATCTCTTCAGGCCACGTTACCCCATGACCTCTCACGA 1506
QY 1871 gacctctgacgaactggaatacttccagagtcaggagattccaggttgaatacaag 1930
|||||
Db 1507 GACCTCTGAGCAACTGGAATCTTCCAGAGTCCAGGGATTTCCAGTTGAATACAAG 1566
QY 1931 acttcccccaaaaacaagaagcgaatttgtatctctatcaattgtctcctcagccac 1990
|||||
Db 1567 ACTTCCCCAAACACACAGACGAATTTGTATCTCTTATCAATTGCTCCCTCAGCCAC 1626
QY 1991 ctctgatacagcatggtatcggcgaagatgtgtgagtcctgcacatgatatgctgcgtga 2050
|||||
Db 1627 CTCTGATCAGCCATGGTATGGGCAAGGATGTGAGTCTGCCATGATATGGCTGCGCTGA 1686
QY 2051 acatcttaagttgtctgtcgaagttggaaccaagaatcacagagatgccaagaacagaa 2110
|||||
Db 1687 ACATCTTAAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAGGAA 1746
QY 2111 acggaaccaatgtctgtgtggtgaggtgtcgaacctttctggccatgaacctataaa 2170
|||||
Db 1747 ACGGACCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGGCCATGAACATTATATAA 1806
QY 2171 atcccaacatatatactgnaaataactgtcttgnaaaatttgaatttctgatacc 2230
|||||
Db 1807 ATCCCAACATATATACTGAAATACTGAAACTGCTTTGAAAAATTGGAATTTCTGATACC 1866
QY 2231 tccagttggccgagagacacggtgtgtaaagatgtgggcaagcagcaggaagaacaacag 2290
|||||
Db 1867 TCCAGTGGGCCGAGAGACACGCTGGGTTAAAGGATGTGGGCAGCAGCAGGGGACACAACAG 1926
QY 2291 aaacacaaggaagcggtgtgtggccggtgtgactgtctggggttgtgtgattggccact 2350
|||||
Db 1927 AAACACAAGGAGGCGGTGTGGCCGGCTGACTGTGCTGGGTTTGTGTGTGGCCACT 1986
QY 2351 cgttgacctggtgcgttccctacgcaatagcagctgcctgtgggaagaaggggtgcccagc 2410
|||||
Db 1987 CGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCTGTGGGGAAGAGGCGTGCACCAGC 2046
QY 2411 cagctggtcttcccggaaccacgcagatccacacctgggcaacctccgtgtttgtctt 2470
|||||
Db 2047 CAGCTGGTCTCTCCCGGACACACGAGATCCACACCCCTGGGCACCTCCGTTTGGTCTT 2106
QY 2471 tttttccctgtgtgaaagaagaacgycacgaccccccttctcaagctggtcactcaga 2530
|||||
Db 2107 TTTTTCCTCCCTGTGTGAAGAAGAAACGGCACGACCCCTTCTCAAGCTGGCTCAGTCAGA 2166
QY 2531 cacattggtgacaaaccctgacagccatgcccagagagaggcctltgaccggccccagagc 2590

Db 2167 CACATTTGGACAAACCCCTGGACAGCCATGCCAGAGAGAGGCCCTTTGACCGGCCACAGAGC 2226
QY 2591 taaagcaccagagaaatcaaatgtcttcctactcagcgltgaccacatttctagtgtg 2650
|||||
Db 2227 TAAAGCACCCAGAGAAATCAATAGCTTCTTACTACGCGTGAACCAACTTTCTAGTGTG 2286
QY 2651 ccagggcccccaaccactcctctgcaagtaccacacccaaccaactgtcttcttccaacag 2710
|||||
Db 2287 CCACGGCCCCACCACTCCTGAGTACCCACACCATCACCTGCTTCTCTTCCAACAG 2346
QY 2711 tgatctgatacttagtttcaattatlttctttagatgatatgacactataataaattt 2770
|||||
Db 2347 TGATCTGATTTCTTAGTTCAATTAATTTCTTTTGATGATATGACACTATATAAATTTT 2406
QY 2771 cattgagaatttcccaattgtatactagttlaaatagcacaglttggaaactgtctgaga 2830
|||||
Db 2407 CATTTGAGAAATTTCTCAATTTGTATCTAGTTAAATAGCACAGTTTGGAAACTTGTGAGA 2466
QY 2831 ctgacttatacataatctaaccgacaagaatcataatcatatgtgtatgtgttagacatt 2890
|||||
Db 2467 CTGACTTTATCAATATATCAACCGACAAAGATCATATCATGTGTATGTGTAGACATT 2526
QY 2891 ttatttcatltgactaaccaggacaglttccagtgatgcaaatgtgtgcctctgtctc 2950
|||||
Db 2527 TTTATTTTCAATTGACTAACCCAGACAGTTTCAGTGATGCAAAATGTGTGCCCTGTGTTG 2586
QY 2951 agctgaacacagtcctcgtgaacttccaaaaaaccttgaaataagtcctccacagltgtataaatt 3010
|||||
Db 2587 AGCTGAACACAGTCCCTGACCTTTCAAAAACCTTGAATAGCTCTCCACAGTGTATAAATT 2646
QY 3011 ggaacaattaggaatttlaaactttagatgatacatgtgttccatttlatltaattt 3070
|||||
Db 2647 GGACAAATTTAGGAATTTTAACTTTAGATGATCATTTGGTTCCATTTTATTTTCAATTTT 2706
QY 3071 atttltglaaltgcaaacaggaactlaaagaacttgaatctcgtlttlaagaattataa 3130
|||||
Db 2707 ATTTTGTAAATGCAAAACAGGACTTAAATGAACCTTGAATCTCTGTTTAAAGATTATTAA 2766
QY 3131 aaacattgtatctatacatatggtccttgaaggaacttccactacatacagga 3190
|||||
Db 2767 AAAACATTTGTATCTATATACATATGGCTCTTGAGGACTTAGCTTTCACATACACTACAGA 2826
QY 3191 tatgatctccatgtagtcacataaaacctgcagagtgatlttccagagtgctcgatactg 3250
|||||
Db 2827 TATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGATTTCCAGAGTGCCTCATACTG 2886
QY 3251 ttaattacatctccatlaggctgaaagaatgaacctagcttctgtatcacagctgtgt 3310
|||||
Db 2887 TTAATTAATCTCCATTAGGGCTGAAGAAATGACCTTAGCTTCTGTATACAGCTGTGTT 2946
QY 3311 gcttltgattgtgttactgtacacagaagtgtgtcactgaggctcgtcgtgtgtcc 3370
|||||
Db 2947 GCTTTGTATGTGTGTACTGTACACAGAAAGTGTGCACTGAGGCTCTGCGTGTGCTCC 3006
QY 3371 gtagtgaanaacctgtgtagccctgcgagltgaagtactgttccatcatgttctaacgtgg 3430
|||||
Db 3007 GTATGGAANAACCTGTAGCCCTGCGAGTTAAGTACTGCTTCCATTTGTTTACGCTGG 3066
QY 3431 aatttcttcccccatggaatgttaagtaaaacttaagtgttgtcatcataataatgttaat 3490
|||||
Db 3067 AATTTTCTCCCATGGAATGTATAAGTAAACTTAAGTGTGTGTCATCAATATAATGTAAT 3126
QY 3491 actaaaaaanaaaaaa 3506
|||||
Db 3127 ACTAAAAAANAANA 3142

RESULT 5
AK025519 3194 bp mRNA PRI 29-SEP-2000
LOCUS AK025519 3194 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21866 fis, clone HEP02379, highly similar to
AF061938 Homo sapiens stauufen protein (STAU) mRNA.

QY 3061 ttccattttattttgtttaatgcaaacaggaacttaaatgaactttgatctctgttttaa 3120
|||||
Db 2772 TTTTCATTTTATTTTGTATGCAACACGAGCACTTAATGAACCTTGATCTCTGTTTAA 2831
QY 3121 agattatataaaacattgtgtatctatataatgctcttgaggacttaagcttccacta 3180
|||||
Db 2832 AGATTATTAAAAACATTGTGTATCTATACATATGAGCTCTTGAGGACTTAGCTTTACACTA 2891
QY 3181 caactacagatatgatccatccatgtagtccataaaccctgcagagtgatltccagagt 3240
|||||
Db 2892 CACTACAGGATATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGATTTCCAGAGTG 2951
QY 3241 ctcgatactgttaattacatctccattagaggctgaaaagaatgacctacgtttctgtata 3300
|||||
Db 2952 CTCGATACTGTATATACATCTCCATTAGGCGCTGAAGAAAGATGACCTACGTTCTGTATA 3011
QY 3301 cagctgtgttgcctttgatgtgtgttactgtacacagaaagtgtgtgcaactgaggtctg 3360
|||||
Db 3012 CAGCTGTGTGCTTTGATGTGTGTACTGTACACAGAGAGTGTGTGCACTGAGGCTCTG 3071
QY 3361 cgtgtgtccgtatggaacacccctgtagccctgcgagtttaagtlactgtcttccattctg 3420
|||||
Db 3072 CGTGTGCTCCGTATGCAAAACCTGTAGCCCTGCGAGTTAAGTACTGCTTCATTCAATTG 3131
QY 3421 ttacgctggaattttctcccatggaatgaagtaaaacttaagttgttcatcaat 3480
|||||
Db 3132 TTTCACGCTGGAATTTTCTCTCCCATGGAATGTAAGTAAACTTAAAGTGTGTGCATCAAT 3191
QY 3481 aaatgtaatactaaaaaanaaaaaa 3506
|||||
Db 3192 AAATGGTAATACTAAAAAANAANA 3217

RESULT 4
AF061940 3142 bp mRNA PRI 04-MAR-1999
LOCUS
DEFINITION Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061940
VERSION AF061940.1 GI:4335950
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3142)
Wickham, L., Duchaine, T., Luo, M., Nabl, I.R. and DesGroselliers, L.
Mammalian stauflen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3142)
AUTHORS DesGroselliers, L., Wickham, L. and Luo, M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1. 3142
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. 3142
/gene="STAU"
288. 1778
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="stauflen protein"
/protein_id="AAD17532.1"
/db_xref="GI:4335951"
CDS
gene

/translation="MKLGKKRMYKPVDPYSRMOSTYNNMRGGAYPPRYFYPPFPVPL
LYQVELSVGGQOFNGKGTROAKHDAAKALRILONEPLPERLEVNGRESEENLNK
SEISOVEEIALKRNLPVNFEVARESGPHMKNFVYKVSVEFVGESEKSKISKNA
AIAVELELKKLPPLPAVERKVPRIKKTKPIVKPQTSPEYGOGINPISRLAQIQAKK
EKEPEYLLTERGLPRRREFVMQVKNHTAEGTGNKKVAKRNAEENMLEIGFV
QROPTKPAKSEKTPIKKPGDGRKVTFFEPGSGDENGTSNKEDEFMYLSHQOLPA
GLPMYDEVAQAVSGGHHKDEFTRAAPNPAKATYTAIARELLYGTSPTAETILK
NNISSHVPHPGLTRPSEQLYLSRVQGFQVEYKDFPKNNKNEFVSLINCOSSQPLIS
HGIGKDYVESHDMALNLIKLSLELDQOSTEMPRGNGPMSVCGRC"

BASE COUNT 862 a 778 c 733 g 769 t
ORIGIN
Query Match 82.6%; Score 2896; DB 10; Length 3142;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 611 cagaaagcataaccctactgtagaactaaatgcaactgtgcatgaaacttggaanaaac 670
Db 247 CAGAAAGCATAAACCCCTACTGTAGAACATAATGCACTGTGCATGAACCTTGGAANAAC 306
QY 671 caatgtataagcctgttgaacccttactctcgatgacgtccaccctataactacaatga 730
Db 307 CAATGTATAAGCCTGTGACCTTACTCTCGGATGACGTCCACCTATTAACATAACATGA 366
QY 731 gaagagtgctatcccccgagtgacttltaccatltccagtlccacccttactatc 790
Db 367 GAGGAGGTGCTTATCCCGGAGGTACTTWTACCATTTCAGTTCACCTTTACTTATC 426
QY 791 aagtggaacttctgtggaagcagcaactttaaagcagaaggaagacaagcagctg 850
Db 427 AAGTGGAACCTTCTGTGGAGGACAGCAATTTAATGGCAAGGAAGACAAGACAGGCTG 486
QY 851 cgaacacgatgtctgtgccaagcggttgagatccctgcagaatgagccctgcagaga 910
Db 487 CGAAACACGATGCTGTGCAAAAGCGTTGAGGATCTCGAAGATGAGCCCTGCCAGAGA 546
QY 911 ggtctggaagtgatgagaagaatccgaagaagaaatctcaataatctgaataagtc 970
Db 547 GGCTGGAGGTGAATGGAAGAATCCGAAGAAGAAATCTCAATAATCTGAATAATAGTC 606
QY 971 aagtgttgagattgcaactaaacgaacttgcctgtgaattcgaagtggccggaga 1030
Db 607 AAGTCTTTGAGATTGCACTTAAACGGAACCTGCTGTGAATTTGAGGTGGCCGGAGA 666
QY 1031 gtggcccaaccacatgaagaacttltgaccaaggttccglttgaggaggttggggg 1090
Db 667 GTGGCCCAACCCACATGAAGAACTTTGTGACCAAGGTTTCGTTGGGAGTTGTGGGG 726
QY 1091 aagtggaaggaagcaagaagatttcaagaanaaatgcgcgcatagtcttcttggg 1150
Db 727 AAGTGGAAGGGAAGCAAGAGATTTCAAAGAAATAATGCCGCCATAGCTGTTCTTGAAG 786
QY 1151 agctgaagaagttaccgcccctgcctgcagttgaacgagtaagaagcctagaatcaaaaaga 1210
Db 787 AGCTGAAGAAGTTACCGCCCTGCTGCAAGTTGAACGAGTAAAGCCTAGAAATCAAAAAAGA 846
QY 1211 aaacaaaaccatagtcagcacaagcacaagcccagaatatagccagggatcaatccga 1270
Db 847 AAACAAAACCCATAGTCAAGCCACAGACAAGCCCAAGATATGGCCAGGGATCAATCCGA 906
QY 1271 ttagccgactgcccagatccagcagcgaaggaaggaagagcagagttacacgctcc 1330
Db 907 TTAGCCGACTGGCCAGATCCAGACGCAAAAAAGAGAAGAGAGCCAGAGTACACGCTCC 966
QY 1331 tcacagagcagagcctcccgccgagggaggtttgtgatgcaggtgaaggttggaaacc 1390
Db 967 TCACAGAGCGAGGCTCCCGCGCCGAGGAGGTTGTGATGCAGGTGAAGTTGGAAGACC 1026
QY 1391 acactgcagaaggaagcagcacaacaagaaggttggccaagcgcaatgcagccgagaaca 1450
Db 1027 ACACGTCAAGAAGAACGGGCAACAACAAGAAGGTGGCCAAAGCGCAATGCAAGCCGAGAACA 1086

Db 552 AGACAGGCTGCGAAACACGATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCC 611

Qy 901 ctgccagagagctgtgagtgtaatggaagaaatccgaagaagaaaaatctcaataatct 960

Db 612 CTGCCAGAGAGGCTGGAGGTGAATGGAAGAAATCCGAAGAAGAAAAATCTCAATAAATCT 671

Qy 961 gaataagtcgaagtgtttgagatgtcacttaacggaacttgctgtgaatttcgagtg 1020

Db 672 GAAATAAGTCAAGTGTGTGAGATGCCACTTAAACGGAACTTGCCGTGAATTTGAGAGTG 731

Qy 1021 gcccgagagagtgcccccacatagaaacttgtgaaccaaggttccggtgggag 1080

Db 732 GCCCGGAGAGTGGCCACCCTGCACATGAAGAACTTTGTGACCAGGTTTCGGTTGGGGAG 791

Qy 1081 ttctgggggaaagtgaaaggaaaagcaagaagattcaaaagaaaaatgcccataagct 1140

Db 792 TTTGTGGGGAAAGGTGAAGGGAAAGCAAGAGATTTCAAAGAAAAATGCCCGCATAGCT 851

Qy 1141 gttcttgagagctgaaagaattaccgccccctgcccgtgcagtgaacgagtaaaagcctaga 1200

Db 852 GTTCTTGAGAGCTGAAGAAGTTACCGCCCTGCCCTGCAGTTGAACGAGTAAAGCCTAGA 911

Qy 1201 atcaaaaaaacaacaaacccatagtcaaagccaacagacaagcccagaatatgtgccaggg 1260

Db 912 ATCAAAAAGAAAAAACCCATAGTCAAGCCACAGACACACCCAGAATATGCCCCAGGGG 971

Qy 1261 atcaatccgattagccgactggtcccagatccagcagcnaaaaaaggagaagagccagag 1320

Db 972 ATCAATCCGATTAGCCGACTGGCCAGATCCAGACAGGCAAAAAAGAGAGAGCCAGAG 1031

Qy 1321 tacacgtcctcaacagagcgagcctcccgcgccgagaggttgtgatgcagtgtaag 1380

Db 1032 TACACGCTCCTCACAGAGCGAGGCTCCCGCGCCGACAGGAGTTTGATGCAGGTGAAG 1091

Qy 1381 gttggaaccacactgcagaagaaacggtcaccaacaagaaggtggccaagcgaatgca 1440

Db 1092 GTTGGAAACCACTGCAGAAAGAAAGGGCACCAACAAGAAAGTTGGCCAAGCGCAATGCA 1151

Qy 1441 gccgagacatgctgtgagatccttgtttcaaaagtcccagcgcgaccccaaaccc 1500

Db 1152 GCCGAGACATGCTGGAGATCCTTGCTTTCAAAGTCCCGCAGCGGCAGCCACCAAAACC 1211

Qy 1501 gcactcaagtcagagagaagaacccataaagaaaacaggggagtgyaagaaaaagtaacc 1560

Db 1212 GCACCTCAAGTCAGAGGAGAGAACACCCATTAAGAAACCAGGGGATGGAAGAAAGTAACC 1271

Qy 1561 tttttgaaacctggtcctgtggatgaaatgtggactagtaataaagaagatgattcagg 1620

Db 1272 TTTTGTGAACCTGGCTCTGGGGATGAATAAGGACTAGTAATAAAGAGGATGAGTTCAGG 1331

Qy 1621 atgacctataagtcacacagcgtgcctgtcgtgaatcttcccatggtgcgcgaggtc 1680

Db 1332 ATGCTTATCTTAAGTCATCAGCAGCTGCCCTGCTGGAATCTTCCCATGGTGCCCGAGGTC 1391

Qy 1681 gcccaagctgtagagtgtaagtcgaagacatcacaccaaagatttacagggcagctccg 1740

Db 1392 GCCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGATTTTACAGGGCAGCTCCG 1451

Qy 1741 aatcctgccaagggccaacggttaactgcatgtagaccggagagttgttgtatgtgggcacc 1800

Db 1452 AATCTGCCCAAGGCCACGGTACTGCCATGATAGCCCCGAGAGTTGTTGATGGGGCACC 1511

Qy 1801 tcgcccacagccgagaccattttaagaataacatccttcaggccaagtaaccacatgga 1860

Db 1512 TCGCCCAAGCCGAGACCATTTTAAAGAAATAACATCTTTCAGGCCACGTAACCCATGGA 1571

Qy 1861 cctctcagagaccctctgagcaactgactatccttcagaggtccagagattccaggtt 1920

Db 1572 CCTCTCAGAGACCCTCTGAGCAACTGGACTATCTTCCAGAGTCCAGGATTCAGGTT 1631

Qy 1921 gaatacaaaacttccccaaaaacaagaagaaattgttatctctatacaattgtccc 1980

Db 1632 GAATACAAGACTTCCCCAAAAACAACAAGACGAATTTGTATCTCTTATCAATTGCTCC 1691

Qy 1981 tctcagccacctctgatcagccatggtatcggcaagatgtgaggtcctgcgatgatg 2040

Db 1692 TCTCAGCCACTCTGATCAGCCATGTTTCGGCAAGAGATGTGAGACTCCTGCCATGATATG 1751

Qy 2041 gctgcgtgaaacatcttaaaagtgcgtctgtgagltggaaccaaaagtlacagagatgcca 2100

Db 1752 GCTGCGCTGAACATCTTAAAGTTGCTGTGTGAGTTGGACCAACAAAGTACAGAGATGCCA 1811

Qy 2101 agaacagaaacggaaccaatgtctgtgtgtggaggtgctgaacctttctgccaatgaa 2160

Db 1812 AGAACAGGAACGGACCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGGCCATGAA 1871

Qy 2161 ccattataaaatcccaacatatatactgaaataactgaaactgtttgaaaatttggaa 2220

Db 1872 CCATTATAAATCCCAACATATATACGAAATACTGAAACTGCTTGAATAATTGGAAT 1931

Qy 2221 ttctgatacctccagtggtggccgagagacacggtgtgtaaagatgtgtgcagcagaggg 2280

Db 1932 TTCTGATACCTTCCAGTGGGCCGAGACACGGTGTTAAAGGATGTGGGCAGCAGCAGGG 1991

Qy 2281 aagacaacagaaacacaagagcggtgtgtgcccgtgactgtgtgtgtgtgtgt 2340

Db 1992 AAGACAACAGAAACACAAGAGGCGGCTGTGGCCGCTGACATGTGCTGGGTTGTTGT 2051

Qy 2341 gatggccactcgtgacctgtgcgtccctacgcaatagcagtcctgtgtgggaagag 2400

Db 2052 GATGGCCACTGCGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCCGTGGGAAGAGG 2111

Qy 2401 gctgccagccagctgttctccggggacaccagagatccacacccctggtgcacctccgt 2460

Db 2112 GCTGCCACGACGCTGTTCTCCCGGGACACACAGAGATCCACACCTGGGCACCTCCGT 2171

Qy 2461 gtttgtcttttttccctgtgtgaaagaagaacggtgcacgtccctctcgaagctgg 2520

Db 2172 GTTGTGCTTTTTTTTCCCTGTGTGAAGAAGAAACGGCAGCACCCTCTCTCAAGCTGG 2231

Qy 2521 ctcactcagaacatltggaacaaacccctggaacagccatgccagagagaggtccttgaacg 2580

Db 2232 CTCACCTCAGACATTTGGGACAAACCTGGACAGCCATGCCAGAGAGAGGCTTTGACCG 2291

Qy 2581 gccccagagctaaaaagcaccagagaaaatcaaatgtctcctactcagcgltgaaccaact 2640

Db 2292 GCCCCAGAGCTTAAAGCACACAGAGAAATCAAAATGCTTCTACTCAGCGTGACCCAACTT 2351

Qy 2641 ttctagtgtgccacggtccccacacactcctgcagtaccacacacatcaccaactgttctc 2700

Db 2352 TTCTAGTGTGCCACGGCCCCACACCTCCTGCAGTACCACACCATCACCACCTGCTTCT 2411

Qy 2701 ctccaacagtatctgtatctctagtttcaattatcttcttgttatgtatatgacatat 2760

Db 2412 CTTCCAACAGTGAATCTGTATTTCTTGTTCATTATTTCTTTTGATTGATATGACACTAT 2471

Qy 2761 ataaaatttcaatttgagaatttctcaaatgtlatctagltaaatagcacaglttgyaaac 2820

Db 2472 ATAAAAATTTCAATTTGAGAAATTTCTCAATGTATCTAGTTAAATAGCACAGTTTGAAC 2531

Qy 2821 ttgtctgagactgaacttatacaataatctaaccgacaagaatcalatcagtgtatgtg 2880

Db 2532 TTGCTGAGACTGAACCTTATCAATAATCTTAACCGACAAAGATCATATCATGTGTATGTG 2591

Qy 2881 gttagacattttatatttcaatttgactaaaccagagacagtttcagtgtgcaaatgtgtgc 2940

Db 2592 GTTAGACATTTTATTTATTTGACTGACTAACCCAGACAGATTTCAGTGAATAATGTGTGC 2651

Qy 2941 cctctgttcagctgaaacagtcctctgacttccaacaccttgaataagtcctccacagt 3000

Db 2652 CCTCTGTTCAAGCTGAACAGAGTCCCTGGACTTTCAAAAACCTTGAATAAGTCTCCACAGT 2711

Qy 3001 tgtataaatggaacaatttaggaattttaaactttagatgatcatlttggtccatttta 3060

Db 2712 TGTATAAATGGACAATTTAGGAATTTTAAACCTTAGATGATCATTTGGTTCCATTTTAA 2771

QY 3061 ttcatatttattttgttaatgcaacagacttaaatgaacttgatctctgtttaa 3120
|||||
Db 2904 TTTCATTTTATTTTGTATGCAACAGGACTTAAATGAACCTTGATCTCTGTTAA 2963
QY 3121 agattattaaaaacatltgtatctatacatatgctcttgaggacttagcttccacta 3180
|||||
Db 2964 AGATTATTAATAAACATGTGTATCTATACATATGCGCTCTTGAGACTTAGCTTTCACATA 3023
QY 3181 cactacagatatgatctccatgtatgccatataaaccctgcagagtgatlttccagagt 3240
|||||
Db 3024 CACTACAGGATATGATCTCCATGTAGTCCATATTAACCTGCAGAGTGATTTTCCAGAGTG 3083
QY 3241 ctcgatactgtaattacatctccatgaaggcctgaagaagaagacctaagcttctgtata 3300
|||||
Db 3084 CTCGATACCTGTTAATTAATCATCTCCATTAAGGCGCTGAAGAAGATGACCTACGTTTCTGTATA 3143
QY 3301 cagctgtgtgcttttgatgtgtgttactgtacacagaagtgtgtgcaactgaagctctg 3360
|||||
Db 3144 CAGCTGTGTGCTTTTGTATGTTGTGTACTGTACACAGAAGTGTGTCACTGAGGCTCTG 3203
QY 3361 cgtgtgtccgatatggaacacctgtgtagccctgcagagtaagtactgttccatctatg 3420
|||||
Db 3204 CGTGTGTCCTGATGGAACCTGTGAGCCCTGCGAGTTAAGTACTGCTTCCATTCATG 3263
QY 3421 ttacgctggaattttctcccatggaatgtaagtaaaccttaagtgttgcataaat 3480
|||||
Db 3264 TTTCAGCTGGAATTTTCTCTCCCATGGAATGTAAGTAAACCTTAAGTGTGTGCATCAAT 3323
QY 3481 aaatgtaataactaaaaa 3506
|||||
Db 3324 AAATGGTAATACTAATAAAAAA 3349

RESULT 3

LOCUS AF061938

DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.

ACCESSION AF061938

VERSION AF061938.1

KEYWORDS GI:4335946

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3217)

AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and Desgroselliers,L.

TITLE Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum

JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)

MEDLINE 99147057

REFERENCE 2 (bases 1 to 3217)

AUTHORS Desgroselliers,L., Wickham,L. and Luo,M.

TITLE Direct Submission

JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900

Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada

Location/Qualifiers

1. 3217

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/map="20q13.1"

1. 3217

/gene="STAU"

363. 1853

/gene="STAU"

/function="RNA binding protein"

/note="alternatively spliced transcript"

/codon_start=1

/product="staufen protein"

/protein_id="AAD17530.1"

/db_xref="GI:4335947"

/translation="MKLGKKPMYKVPDYSRMGSYYNYNMRGAYPRRYFPVPPL

LYVELSVGGQOFNGKKTROAKHDAAKALRILONEPLPERLEVNGRESEENLINK
SEISOVEIALKRLNLPVNEVARESGPPHAKNEVTKVSVEFVGESEKSKISKNA
AIAVLEELKLPPLPAVERVKPRIKKTKPIVKPQTSPEYGOGILSRLOIOAKK
EKEPEYLLTERGLPRRREFVMQKVGNHTAEGTGTKKVAKRNAENMLEILGFKVP
QROPTPALKSEKTPIKKPGDGKVTFFEPGSDENGTSNKEDEFMPYLSHOQLPA
GILPMVPEVAOAVGSQGHHTKDFTRAAPNPAKATVTAMIAARELLYSGTSPAETILK
NNISSGHVPHGPLTRPSBQLDYSRVQGFQVEYKDFPKNNKNEFVSLINCSSQPLIS
HGIGKDVESCHDMAALNLIKLSLELDQOSTEMPTGTNGPMsvcgrc"

Query Match 83.1%; Score 2914.8; DB 10; Length 3217;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 3215; Conservative 0; Mismatches 2; Indels 289; Gaps 1;

QY 1 acttcctgcggcgtgcggcgcccttcacgcgtcttcacgcgttcgcgcgcgtgcgcgt 60
|||||
Db 1 ACTTCCTGCCGGCTGCGGGCGCCTGAGCGCTTTCAGCGTTTGCGGGCGGCGCGCTG 60
QY 61 ctctcgcgtccgcgtcttccttcacgcgtcttcacgcgtcttcacgcgtcttcacgcgt 120
|||||
Db 61 CTCTCTCGGCTCCGCTTCTCTTGACCGCTCCCGCCCGCGCGCGCGCGCGCTC 120
QY 121 ctccacgcccactccgcgtcttccttcacgcgtcttcacgcgtcttcacgcgtcttcct 180
|||||
Db 121 CTCACGCGCCACTCCGCTCTTCCCTCCCTTCGTCCTTCCTCTCCCTTTTTCCTT 180
QY 181 ctccctccctccctccgc 240
|||||
Db 181 CTTCCTTCCCTCTCTCGCGCCACCGCCAGGACCGCGCGCGCGCGGAGAGTCCGAGC 240
QY 241 agcagccagagtttaataaccacttaacctcagaactgaacaagaacaacatgttcc 300
|||||
Db 241 AGCAGCCAGAGTTTATTAACCACTTAACCTTCAGAACTGAACAAGACAACATTTGTTCC 300
QY 301 tggaaagccctctttaaagaagtagaactttagacttcataagcaactgaactgaact 360
|||||
Db 301 TGGAAAGCCCTCTTTTAAAAA----- 322
QY 361 cactgaagctgtttacctgcatlttgtcaactttgttgaagtgaccatgtctcaagtc 420
|||||
Db 323 ----- 322
QY 421 caagtgaagttcagaaccatctgtctctcagggagccaataactgaacaagaac 480
|||||
Db 323 ----- 322
QY 481 cagtcctctcacaagcctttagatgattctccttactactagctctctgcctct 540
|||||
Db 323 ----- 322
QY 541 gaaatgcagtagaccatccaactcgtcttaccctctgcatctataatccaacc 600
|||||
Db 323 ----- 322
QY 601 agtcagctgcagaaagcataaccctactgtagaactaaatgaactgtgcatgaact 660
|||||
Db 323 -----AGAAAGCATATAACCCCTACTGTAGAATAAATGCACCTGTGCATGAACCTT 371
QY 661 ggaaaaaaaccaatgtataagcctgttgacccttactctcggatgagtcacactataac 720
|||||
Db 372 GGAATAAACCAATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATAC 431
QY 721 tacaacatgagagaggtgtctatcccccgaggtacttltaccatltccagttccacct 780
|||||
Db 432 TACAACATGAGAGAGGTGCTTATCCCGAGGAGTACTTTTACCACTTTCAGTTCCACCT 491
QY 781 ttactttacaatggaacttctgtggaagagacagcaatttaatggaagaagaaga 840
|||||
Db 492 TTACTTTATCAAGTGAACCTTCTGTGGAGGAGACAGCAATTTAATGGCAAGAAAGACA 551
QY 841 agacagctgcgaaacacgatgtctgtccaaagcgttgaggtatccttgcaaatgagccc 900

Db 684 AGACAGGCTGCCAACACGATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGATGAGCCC 743

QY 901 ctgccagagaggtctgtagtgaatyggaagagaatccgaagaagaataatctcaataatct 960

Db 744 CTGCCAGAGAGGCTGGAGGTGAATGGAAGAGAAATCCGAGAAGAAAATCTCAATAAATCT 803

QY 961 gaaataagtcaagtgtttgagattgcacttaaaacgaacttgctgtgaatttcgaagtgtg 1020

Db 804 GAAATTAAGTCAAGTGTGTGAGATTGCACCTTAAACGGAACTTGCTGTGAATTTCGAGGTG 863

QY 1021 gcccgagagagtgccccaccacatagaaagaactltgtgaccgaagtlttcggttgggag 1080

Db 864 GCCCGGAGAGTGGCCACCCACATGAAGAACTTTGTGACCAAGGTTCGGTTGGGGAG 923

QY 1081 ttgtgaggggaagtgaaggggaaagaagaagatttcaagaanaaatgcccagctagct 1140

Db 924 TTTGTGGGGGAAGGTGAAGGGAAAGCAAGAGATTTCAAAGAAAAATGCCGCCATAGCT 983

QY 1141 gttcttgagagagctgaaagaagttaacggccctgctgcagttgaacgagtaaacctaga 1200

Db 984 GTTCTTGAGGAGCTGAAGAAATTACCGCCCTGCTGCTGAGTTGAACGAGTAAAGCCTAGA 1043

QY 1201 atcaaaaaaagaaaaacaaaccatagttcaagccacagagaagccagaatatgccaaggg 1260

Db 1044 ATCAAAAAGAAAACAAAACCCATAGTCAAGCCACAGACAGCCAGAAATATGGCCAGGGG 1103

QY 1261 atcaatccgaattagccgactgcccagatccagcagcgcaaaaaagagaagagccagag 1320

Db 1104 ATCAATCCGATTAGCCGACTGGCCACAGATCCAGCAGCGCAAAAAAGAGAAGAGCCAGAG 1163

QY 1321 tacagctcctcacagagcgagcgctcccgcgccgcaaggagtttgtatgcaggtgaag 1380

Db 1164 TACACGCTCCTCACAGAGCGAGCCCTCCCGCGCCGAGGGAGTTGTGATGCAGGTGAAG 1223

QY 1381 gttggaaccacactgctgcaagygaaacygcaaccaagaaggttggccaagcgcaatgca 1440

Db 1224 GTTGGAAACCCACACTGCAGAAAGAACGGGCGACCAACAAGGTTGCCAAGCCCAATGCA 1283

QY 1441 gccgagaaacatgctgtagatccttggtttccaaagtcgccgagcgcgagcccaaaccc 1500

Db 1284 GCCGAGAACATGCTGGAGATCCTGGTTTCAAAAGTCCCGCAGCGGACCCACCAAAACC 1343

QY 1501 gcactcaagtcagagagagaagacacccataaagaaccaaaggga tgyaagaaagtaacc 1560

Db 1344 GCACCTCAAGTCAGAGAGAGAACCCCATAAAGAAACAGGGGATGGAAGAAAGTAACC 1403

QY 1561 tttttgaaacctgctcgtggyga tgaaaatggyactagtaataagagatgagttcag 1620

Db 1404 TTTTGTGAACCTGGCTCTGGGATGAAATGGGACTAGTAATAAGAGGATGAGTTCAGG 1463

QY 1621 atgcctatactaagtcatacagcagctgcctgctggaattctcccatgtgtcccgagtc 1680

Db 1464 ATGCCTTATCTAAGTCATCAGCAGCTGCCCTGCTGGAATTCTTCCCATGTGCCCGGAGGTC 1523

QY 1681 gcccaagctgtagagttagtcaaagacatcacaccaagaatttaccagggcagctcgg 1740

Db 1524 GCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGATTTTACCAGGGCAGCTCCG 1583

QY 1741 aatcctgccaagggccacgylaactgcatgatagcccagagttgtgtatgggggcacc 1800

Db 1584 AATCCTGCCAAGGCCACGGTAACTGCCATGATAGCCGAGAGTTGTGTATGGGGGCACC 1643

QY 1801 tcgcccacagccgagaccattttaagaataacatctcttcagggccaagtaaccatgga 1860

Db 1644 TCGCCACAGCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCACGTAACCCCATGGA 1703

QY 1861 cctctcaagagaccctctgagcaactygactatcttccagagttccagggattccaggtt 1920

Db 1704 CCTCTCACGAGACCCTCTGAGCAACTGGACTATCTTTCCAGAGTCCAGGGATTTCCAGGTT 1763

QY 1921 gaatacaaaagactcccccaaaaaacaagaagaagaaattgtlatctctlatcaattgctcc 1980

Db 1764 GAATACAAGACTTCCCCCAAAAACAAGAAGAAATTGTATCTCTTATCAATTGCTGCC 1823

QY 1981 tctcagccaaccttgatcagccaatggtlatcggcaagagatgtagtccctgcataatg 2040

Db 1824 TCTCAGCCACCTCTGATCAGCCATGGGTATCGGCAAGGATGTGAGTCTCCATGATATG 1883

QY 2041 gctgcgtgaacatctttaaagttgctgtcgtgagttggaaccaacaagtaacagatgcca 2100

Db 1884 GCTGCGCTGACACATCTTAAAGTTGCTGCTGAGTTGACCAACAAGTACAGAGATGCCA 1943

QY 2101 agaacaggaaacggaaccaatgtctgtgtgtgtggaagtgctgaaccttcttcggcatgaa 2160

Db 1944 AGAACAGGAACGGACCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAA 2003

QY 2161 ccaatlataaaatcccaacataataactgaaaaaactgaaaactgctltgaaatttgyaat 2220

Db 2004 CCATTATTAATCCCAACATATATACTGAAAATACTGAAACTGCTTTGAAAATTTGGAAT 2063

QY 2221 tcttgatacctccagtgggccgagagacaagtggtgtaaagatgtggcagcagcagg 2280

Db 2064 TTTCTGATACCTCCAGTGGGGCCGAGACACGGTGGGTAAAGATGTGGGCAGCAGCGG 2123

QY 2281 aagacaacagaaacacaaaggagcggtgtgtgcccgtggaactgtgtgtggtttgtgt 2340

Db 2124 AAGACAACAGAAACACACAGAGCGGGCTGTGCGCCGGCTGAGCTGTGCGGTGTTGTTGT 2183

QY 2341 gatggccaactcgtgacctgtgcggttccctaagcaatagcagctgcctgtggygaagaag 2400

Db 2184 GATGGCCACTCGGTGACCTGGCGGTTCCTTACGCAATAGCAGCTGCCGTGTGGGAAGAAGG 2243

QY 2401 gctgcccaagccaagctgtgtctcccggygacaaccaagataccacaacctggygacctcgt 2460

Db 2244 GCTGCCACGCGAGCTGGTGTCTCCGGGACACCAGCAGATCCACACCCGTGGGCACTTCGT 2303

QY 2461 gtttggttttttttccctgtgtgaaagaagaacgcaagacccttccaagctgg 2520

Db 2304 GTTTGGTCTTTTTCCTTCCTGTGTGAAGAAGAAGAACGGCACGACCCCTTCTCAAGCTGG 2363

QY 2521 ctcactcagacacattggygacaacacctgycagcgcaatgycagaagagcgcttggaccg 2580

Db 2364 CTCACTCAGACACATTTGGGACAACCCCTGGACAGCCATGCCAGAGAGAGGCCCTTTGACCG 2423

QY 2581 gccccaagactaaaagcaccgaagaanaatcaaatgtctcctaactcagcgttgaccccaact 2640

Db 2424 GCCCCAGAGCTAAAAGCACCAAGAGAAAATCAAAATGCTTCTACTCAGCGGTGACCCAACTT 2483

QY 2641 ttctaagtgycacggccccacacclctcgtcagtaaccacaaccaactgccttctc 2700

Db 2484 TTCTAGTGTGCCACGGCCCAACCACTCTGCAGTACCCACACCATCACCACTGCTTCT 2543

QY 2701 ctccaacagtgatctgtattcttaagtttcatatttcttltgattgatatgacactat 2760

Db 2544 CTTCCAACAGTGATCTGTATCTTAAGTTTCATATTTCTTTGATTGATATGACACTAT 2603

QY 2761 ataaatttcatatttgagaatttctcaattgtlatctagtttaaatagcacagtttgyaac 2820

Db 2604 ATAATAATTTTCATTTGAGAAATTTCTCAATGTATCTAATTAAATAGCACAGTTTGAAC 2663

QY 2821 ttgtctgagaactgacttlatcaataatctaacccgaacaagatcatatccatgltatgtg 2880

Db 2664 TTGTCTGAGACTGACTTTATTCATAATATCTAACCCGACAAAGATCATATTCATGTATGTG 2723

QY 2881 gtagacatttltatltcatgtacttaaccagagacagttcagtgatgcataattgtgtgc 2940

Db 2724 GTTAGACATTTTATTTATTCATGACTAACCCAGGACAGTTTCAGTGATGCAAAATGTGTGC 2783

QY 2941 cctctgtcagctgaacacagctcctggaacttcaaaaaacttgaataagttcccaagct 3000

Db 2784 CCTCTGTTACGCTGAACACAGTCTTGACTTTCAAAAACCTTGAATAAAGTCTCCACAGT 2843

QY 3001 tgtataaatgagacaatttgaagtlttaaaccttagatgatcatattgttccatttla 3060

Db 2844 TGTATTAATTGGACAATTTAGGAATTTTAAACCTTTAGATGATCATTTGTTCCATTTTTA 2903


```
|||||
Db 3061 TTTCATTTTATTTTGTATATGCAACAGGACTTAATGAACCTTGATCTCTGTTTAA 3120
QY 3121 agattataaaaaacattgtgtatcatatacatatgctcttgaggacttagcttcaacta 3180
Db 3121 AGATTATTAATAAACATTGTGTATCTATACATATAGGCTCTTGAGGACTTAGCTTCACTA 3180
QY 3181 cactacagatatgatctccatgtagctccatataaaccctgcagagtatttccagagt 3240
Db 3181 CACTACAGATATGATCTCCATGTAGCCATATATAACCTGCAGAGTATTTCCAGAGTG 3240
QY 3241 ctcgatactgttaatacatatccatcattaggcgctgaagaagatgacctacgttctgtata 3300
Db 3241 CTCGATACGTTAATACATCTCCATAGGGCTGAAGAAGATGACCTACGTTCTGTATA 3300
QY 3301 cagctgtgtctgttttgatgtgtgttactgttacacagaagtgtgtgcaactgagctctg 3360
Db 3301 CAGCTGTGTGCTTTTGATGTGTGTACTGTACACAGAAGTGTGTGCACTGAGGCTCTG 3360
QY 3361 cgtgtgtccgatatgaaaaacctgtgtacccctgcgagttgaagtactgtctccatcattg 3420
Db 3361 CGGTGTGCTCCGTATGGAACCTGTGTAGCCCTGCGAGTTAAGTACTGCTTCCATTCAATTG 3420
QY 3421 ttacagctgaatttctcccatggaatgtaagtaaaactaagttgtgtcatcaat 3480
Db 3421 TTACGCTGGAATTTTCTCCCATGGAATGTAAGTAAACTTAAGTGTGTGTCAATCAAT 3480
QY 3481 aaatgtaataactaaaaa 3506
Db 3481 AAATGTAATACTAAAAA 3506
```

```
RESULT 2
AF061941 3349 bp mRNA PRI 04-MAR-1999
LOCUS Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,
DEFINITION complete cds.
ACCESSION AF061941
VERSION AF061941.1 GI:4335952
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3349)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and Desgroselliers,L.
TITLE Mammalian stauflen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3349)
AUTHORS Desgroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1. 3349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. 3349
/gene="STAU"
495. 1985
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="stauflen protein"
/protein_id="AAD17533.1"
/db_xref="GI:4335953"
/translation="MKLGKPMYKRPVDPYSRMOSTNYNMRGAYPRRYFPVPVPL
LYQVELSVGGQOFNGKGTROAKHDAAKALRILONEPLPERLEVNGRESEENLNK
```

```
SEISOFEIALKRNLPVNEFVARESGPRHKNFVTKVSVEFVEGEGSKISKNA
AIAVLELKKLPPLPAVERVKPRIKKTKPIVKPQTSPEYGGINPISRLAQIOAKK
EKEPEYTLTERGLPRRREFVMQVGNHTAEGTGNKKVAKRNAENMLEILGFKVP
QROPPTKPKLSEKTPIKKPGDGRKVTFFEPGSGDENGTSNKEDFRMPYLSHQLP
GLIPMPEVAAGVSGOHTKDFTRAAPNPAKATVTAMIARELYGTSPTAETILK
NMISSGHPHGPLTRPSEOLDYLSRVQGFQVEYKDFPKNNKNEVSLINCSSQPLIS
HGIGKDVESCHDMAALNLIKLSLSDQGSTEMPRTGNGPMSVCGRC"
BASE COUNT 913 a 838 c 777 g 821 t
ORIGIN
Query Match 84.4%; Score 2958.8; DB 10; Length 3349;
Best Local Similarity 91.98; Pred. No. 0;
Matches 3222; Conservative 0; Mismatches 127; Indels 157; Gaps 3;
QY 1 acttcctgcgggctgcgggcgctgagcgctcttcagcgtttgcgggcgctgcgcgt 60
Db 1 ACTTCCTGGCGGCTGGCGGCGCTGAGCGCTCTTCAAGCGTTTGCGGCGGCGGCGGT 60
QY 61 ctctctcgctcccgctctcctttgacgcgcctcccccccgccggcgcgcgccgcctc 120
Db 61 CTCTCTGGCGCTCCGCTTCTTGTGACCGCCTCCCCCCCCCGGCGGCGGCGGCGCTC 120
QY 121 ctcaagcgcaactccgcctctcctccctctcgctccctctcctctcctcttccctt 180
Db 121 CTCACGGCCACTCCGCTCTTCCCTCTCCCTTCTGTCCTTCTTCTCTCTCTCTCTT 180
QY 181 ctccctccctccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
Db 181 CTTCCTTCCCTCCCTCTGCGCCGCGCACCGCCAGACCGCGCGGGGAGCAGATCCGAGC 240
QY 241 agcagccagagttatataaccacttaacctctcgaactgaacaaagacaacattgtcc 300
Db 241 AGCAGCCAG----- 249
QY 301 tggaaagccctctttaaagaagtagaacttagacttcatactgactgaatlaactg 360
Db 301 TGGAAGCCCTCTTAAAGAGTAGAATTATTAACCACTTAACCTCTCAGAACTGA 413
QY 421 caagtgcaggttcagaaccatctgtctctctcaaggagccaataactgaacaagaac 480
Db 421 CAGTGGCGTGATTCGGCTCAGCTGCAACCTCCACCTCCAGGTCAGGATTTGCCACTT 353
QY 481 cagctctctctccacagcctttagatgattcctctactactactagctctctgcctct 540
Db 481 CAGCTCCCGATAGCTGAGATTACAGAGATTATTAAACCACTTAACCTCTCAGAACTGA 413
QY 541 gaaatgcaggtgagcccatccaactcgtcttaccctctgcatactatcatccacc 600
Db 541 ACAAGACACATGTTCTTGGAAACGCCCTCTTT----- 448
QY 601 agtgcagctgcagaagaacataaccctactgtagaactaaatgcactgtgcatagaact 660
Db 601 -----TAAAAAAGAAAGCATTAACCCCTACTGTAGAATAAATGCACCTGTGCATGAACCTT 503
QY 661 ggaaaaaaaccaatgtataagcgtgtgaacccctactctcgagatgcagtcacataac 720
Db 661 GGAATAAACCAATGTATTAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATTAAC 563
QY 721 tacaacatgagagaggtgtctatcccgaggtactttaccattccagttccact 780
Db 721 TACAACATGAGAGAGGTGCTTATCCCGAGGTACTTTTACCACTTTCCAGTTCACCT 623
QY 781 ttacttataagtggaacttctgtgtgagagacagcaatttaatgcaagaagaca 840
Db 781 TTACTTTATCAAGTGAACCTTCTGTGGGAGGACAGCAATTTAATGGCAAGAAAGACA 683
QY 841 agacagctgcgaacaacagatgctgtctgcacaaagcgttgagatcctgcagaatgagccc 900
Db 841 |||||
```

QY 901 ctgcagagaggctggaagtgaatggaagagaatccgaagaagaanaatctcaataaatct 960
Db 901 CTGCAGAGAGGCTGGAGGTGATGGAAGAGATCCGAGAGAAGAAAATCTCAATAAATCT 960
QY 961 gaataaagtcaagtgtttgagattgcacttaaacgnaacttgccgtgtaattccgagtg 1020
Db 961 GAAATAGTCAAGTGTGTGAGATTGCCACTTAACGGAACTTGCCGTGTAATTGAGGTG 1020
QY 1021 gcccggaagagtggcccaacccacatgaagaactttgtgaccaaggttctggttgggag 1080
Db 1021 GCCCGGAGAGTGGCCACCCACATGAAGAACTTTGTGACCAAGGTTTCGTTGGGGAG 1080
QY 1081 ttgtgggggaaagtgaaagggaaaaaagcaagaagattcagaagaaaaatggcccatagct 1140
Db 1081 TTTGTGGGGGAAAGTGAAGGGGAAAAAGCAAGAGATTTCAAAGAAAAATGCCCATAGCT 1140
QY 1141 gttcttgaggagctgaagaagttaccgcccctgctgcagttgaacgagtaaaagcctaga 1200
Db 1141 GTTCTTGAGGAGCTGAAGAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAGCCTAGA 1200
QY 1201 atcaaaaagaanaacaaacccatagtcaagccacagacaagcccaataatgtgccaggg 1260
Db 1201 ATCAAAAAGAAAACAAACCCATAGTCAAGCCACAGACAGCAAGCCAGATATGGCCAGGG 1260
QY 1261 atcaatccgatttagccgactggcccagatccagcagcaaaaaaggaagagccagag 1320
Db 1261 ATCAATCCGATTAGCCGACTGGCCCCAGATCCAGCAGGCAAAAAAGGAGAGAGCCAGAG 1320
QY 1321 tacacgtcctcacaagaagcgaagcctcccgcgccgcaaggagtttgtatgcagtgtaag 1380
Db 1321 TACACGCTCTCACAGAGCGAGGCTCCCGCGCCGAGGGAGTTTGTATGCAAGTGGAAG 1380
QY 1381 gttggaaccacacactgcagaaagaaacgggcacaaacaagaagtgycgaagcgcaatgca 1440
Db 1381 GTTGGAAACCACTGCAGAGAGAAACGGGCACCAACAAGAAGTGGCCAAAGCCGCAATGCA 1440
QY 1441 gccgagaacatgtctgagatccttgtgttcaaatcccgcaagcgagcccaccaaaccc 1500
Db 1441 GCCGAGACATGTCTGAGATCCTTGGTTTCAAAAGTCCCGCAGCGGCAACCAAAACC 1500
QY 1501 gcactcaagtcatagagaagaacacccataaagaacacagggagtgyaagaanaagtaac 1560
Db 1501 GCACTCAGTCAAGAGAGAGAACCCCATTAAGAALACCAGGGGATGGAAGAAAAGTAACC 1560
QY 1561 tttttgaacctggtcctggygagtgaaaatggygactagtaataaagagatagttcag 1620
Db 1561 TTTTGTGAACCTGGCTCTGGGATGAAATGGGACTAGTAATTAAGAGGATGAGTTCAGG 1620
QY 1621 atgcctatctaagtcatacagcagctgcctgtcgtgnaattctcccatgtgtcccgagtc 1680
Db 1621 ATGCTTATCTTAAGTCATCAGCAGCTGCCTGCTGGAATTTCTCCCATGTGTGCCGAGGTC 1680
QY 1681 gcccaagctgtagagtttagtcaaggacatcacaccaaagattttaccagggcagctccg 1740
Db 1681 GCCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGATTTTACCAGGGCAGCTCCG 1740
QY 1741 aatcctgccaagggccaaggtlaactgcatgatagcccgagagttgtgtatggggcacc 1800
Db 1741 AATCTGCCAAGGCCACGGTAACTGCCATGATAGCCCGAGAGTTGTGTATGGGGGCACC 1800
QY 1801 tcgcccacagccgagaccatttaagaataatacatctctcagggccaqtaacccatgta 1860
Db 1801 TCGCCACAGCCGAGACCATTTAAGAATAACATCTCTCAAGCCACGTACCCCATGGA 1860
QY 1861 cctctcacgagaccctctgagcaactgactatcttccagagttccagagattccaggtt 1920
Db 1861 CCTCTCAGAGAGACCTCTGAGCACTGACTATCTTCCAGAGTCCAGGGATTCAGGTT 1920
QY 1921 gaatacaaaagacttccccaaaaacaacaagaagaaatgttatctctatcaattgtcc 1980
Db 1921 GAATACAAAGACTTCCCCAAAAACAACAAGAACGAATTTGTATCTCTTATCAATTGCTCC 1980

QY 1981 tctcagccaacctctgtatcagccaatggtatcggcaaggaatgtggaagtcctgcacatgatg 2040
Db 1981 TCTCAGCCCACTCTGATCAGGCCATGGTATTCGGCAAGGATGTGAGTCTCTGCCATGATG 2040
QY 2041 gctgcgctgaacatcttaagttgctgtctgagttggaaccaacaagtaacagagatgcc 2100
Db 2041 GCTGCGCTGAACATCTTAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCA 2100
QY 2101 agaacagaaacggaaccaatgtctgtgtgtggaagtgctgaacctttctgscatgaa 2160
Db 2101 AGAACAGGAACGGACCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGGCATGAA 2160
QY 2161 ccatataaaatcccaacatatatactgnaaaaatactgnaaactgctltgnaaatltggaat 2220
Db 2161 CCATTATAAATCCCAACATATATACTGAAATACTGAACACTGCTTGAAATTTGGAAT 2220
QY 2221 ttctgtatactccaagtgycggaagaacacagtggttaaggaatgtgycagcagcaggg 2280
Db 2221 TTCTGATACCTCCAGTGGGGCCGAGACACAGGTGGGTAAAGATGTGGGCAGCAGAGG 2280
QY 2281 aagacaacagaaacaacaagagcgcgctgtgycgcyctggaactgtgtggtgttgtt 2340
Db 2281 AAGACAACAGAAACACAAAGAGGGCGCTGTGGCCGGCTGGACTGTGCTGGGGTTGTGT 2340
QY 2341 gatgccaactcgttgacctgcggtccctacgcaatagcaagctgcctgttggygaagaag 2400
Db 2341 GATGGCCACTCGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCCGTGGGGAGAAGG 2400
QY 2401 gctgccagccaagctgttctcccgggacaccagcagatccacaacctgycgaactccgt 2460
Db 2401 GCTGCCAGCCAGCTGTGTCTCCGGGACACCAAGACAGATCCACACCTGGGCACCTCCGT 2460
QY 2461 gtttgctcttttttcccccgtgtgtaagaagaacgycagcagccctctcgaactg 2520
Db 2461 GTTTGGTCTTTTTCCTCCCTGTGTGAAGAAGAAACGGCACGACCCCTTCTCAAGCTGG 2520
QY 2521 ctcactcagcacacattggygaacaaacccctgycagcagccatgycagaagaagcccttga 2580
Db 2521 CTCACCTCAGACACATTGGGACAAACCTTGACAGCCATGCCAGAGAGAGGCGCTTTGACCG 2580
QY 2581 gccccagactaaaagcaccagaagaaaalcaaatgcttccctactcaagcgtgaaccaact 2640
Db 2581 GCCCCAGAGCTAAAAAGCACACAGAGAAAATCAAAATGCTTCCTACTCAGCGTGACCAACTT 2640
QY 2641 ttctagtgtccaagggccccaacacctcctgcagtacccaacatcaccaactgcttct 2700
Db 2641 TTCTAGTGTGCCACGGCCCCACCACTCCTGCAGTACCCACACCATCACACACTGTTCT 2700
QY 2701 ctccaacagtgaatctgtatctcttagtttcatatttctctttgtatgatgaactat 2760
Db 2701 CTTCCAACAGTGAATCTGTATTTAGTTTCAATTAATTTCTTTTGATGTGATATGACACTAT 2760
QY 2761 ataaaatttcatlttgagaatttctcaattgtaatctagttaaatagcacagtttgyaaac 2820
Db 2761 ATAAAAATTTTCAATTTGAGAAATTTCTCAATTTGTATCTAGTTAAATAGCACAGTTTGAAC 2820
QY 2821 ttgtctgagactgaacttatacaataatctaaccgacaaagatcatatcatgtgtatgt 2880
Db 2821 TTGTCTGAGACTGACTTATCAATAATCTTAACCGACAAGATCATATCCATGTGTATGTG 2880
QY 2881 gttagacatttttatttcatlttgactaaacccagacagtttcagtgatgcaaatgtgtgc 2940
Db 2881 GTTAGACATTTTATTTTCAATTTGACTTAACCCAGGACAGTTTCAGTGATGCAAAATGTGTGC 2940
QY 2941 cctctgtgtcagctgaacacagtcctgtgacttcaaaaaacctgtaataagtcctccacagt 3000
Db 2941 CCTCTGTTCAGCTGAACACAGTCTGTGACTTTCAAAAAACCTTGAATAAGTCTCCACAGT 3000
QY 3001 tgtataaattgacaattttagaattttaaactttagatgatacatlttgttcaattttaa 3060
Db 3001 TGTATAAATTGGACAATTAGGAATTTTAAACTTTAGATGATCATTTGGTTCCATTTTTA 3060
QY 3061 ttcatatttatttttgtttaatgcaaacagagacttaaatgaaacttgaatctctgtttaa 3120

22	150.8	4.3	160979	52	AC055879	AC055879	Homo sapi
23	149.6	4.3	3173	34	AK001576	AK001576	Homo sapi
24	149	4.2	332811	72	AL390198	AL390198	Homo sapi
25	133.4	3.8	194407	54	AC068845	AC068845	Homo sapi
26	125.2	3.6	2008	34	AK023314	AK023314	Homo sapi
27	119	3.4	2847	69	HSRNATTFI	X83973	H. sapiens m
28	110.2	3.1	5237	53	DROSTAUFEN	M69111	D. melanogas
29	110.2	3.1	61267	28	AC004336	AC004336	Drosophil
30	110.2	3.1	120625	38	AC017563	AC017563	Drosophil
31	110.2	3.1	267271	28	AE003800	AE003800	Drosophil
32	105.8	3.0	8409	29	AF225924	AF225924	Drosophil
33	103	2.9	8002	29	AF225925	AF225925	Musca dom
34	85.4	2.4	171591	50	AC027018	AC027018	Homo sapi
35	74	2.1	7218	81	I66494	I66494	Sequence 14
36	68	1.9	7218	81	I66494	I66494	Sequence 14
37	65.4	1.9	60009	48	AC023436	AC023436	Homo sapi
38	63	1.8	180968	62	AC078996	AC078996	Mus muscu
39	63	1.8	258550	35	AC010758	AC010758	Homo sapi
40	62.6	1.8	42099	29	CELF55A4	U67949	Caenorhabdi
41	62.4	1.8	141892	47	AC023197	AC023197	Mus muscu
42	62.2	1.8	89669	69	HSJ340H11	AL109755	Human DNA
43	61.4	1.8	77127	36	AC011801	AC011801	Homo sapi
44	61.2	1.7	92709	37	AC015716	AC015716	Homo sapi
45	61	1.7	80725	34	AL137078	AL137078	Human DNA

ALIGNMENTS

```

TSPEYGGGINPISRLAQIQOAKKEKEPEYTLTERGLPRRREFVMQVKNHTAEGTG
TNKKVAKRNAENMLEILGFVPQROPTKPAKSESEKTPIKKPGDGRKVTFFEPGSGD
ENGTSNKEDEFMRPYLSHOQLPAGILPMWPEVAQAVGSQGHHTKDFTRAPNPAKAT
VTAMIARELLYGGTSPYTAETILKNNISSGHVPHGPLTRPSEQLDYLSRVGQFOVEYKD
EPKNKNKEFVSLINCSSQPLPLISHGIGKDVESCHDMAALNLIKLLSELDOOSTEMPRT
GNGPMSVCGRC"
BASE COUNT      965 a      875 c      790 g      876 t
ORIGIN

Query Match      99.9%; Score 3502.8; DB 10; Length 3506;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3504; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  actctctgcgagctgcgagcgccctagcgcctctcagcgtttgcgcgagctgcgcgt 60
      |||||||
Db      1  ACTTCTGCCGCGGCTGCGGCGCCCTGAGCGCCTTCAAGCGTTGCGCGCGGCGGCTGCGCGT 60

QY      61  ctctctgcgctccgcgtctctctgaccgcctccccccccgcgcgcgcgcgcgcctc 120
      |||||||
Db      61  CTCTCTCGGCTCCCGCTTCTTGAACGCGCTCCCGCCCCCGGCGCGGCGCGCGCTC 120

QY      121  ctcaacgagcactccgcctctctccctcccttcgctccctctctccctctcccttttccct 180
      |||||||

```

```

RESULT      1
LOCUS       AF061939
DEFINITION  AF061939 3506 bp mRNA PRI 04-MAR-1999
            Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
            complete cds.
ACCESSION   AF061939
VERSION     AF061939.1 GI:4335948
KEYWORDS
SOURCE      human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3506)
AUTHORS     Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE       Mammalian staufen is a double-stranded RNA and tubulin binding
            protein which localizes to the rough endoplasmic reticulum
            Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL     99147057
MEDLINE
REFERENCE   2 (bases 1 to 3506)
AUTHORS     DesGroselliers,L., Wickham,L. and Luo,M.
TITLE       Direct Submission
            Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
            Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
            source
            1..3506
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="20"
                /map="20q13.1"
                1..3506
                /gene="STAU"
                409..2142
                /gene="STAU"
                /function="RNA binding protein"
                /note="alternatively spliced transcript"
                /codon_start=1
                /product="staufen protein"
                /protein_id="AAD17531.1"
                /db_xref="GI:4335949"
                /translation="MSOVQVQVQNPASALSGSQLNKNSLLSOPILMSIPTSTSLPSS
                ENAGRPIONSALPSASITSTSAAESITPTVELNALCMKLGKKPMYKPVDPYSRMOST
                YWYNMGAGVAPRYFYPPVPPLLYQVELSVGGQFNGKGTROAKHDAAKALRILL
                ONEPLPERLEVNGRESEENLNKSEISOVEIALKRNLPVNFVARESGPPHMKNEVT
                KSVGEEFVGEGEKSKISKKNAAIAVLEELKRLPLPAVERVKPRIKKTKPIVKPO

```

QY	181	cttccttccccctcctccgcgcgcaaccgccgaagacgcgcgcggcggygagcagatcggagc	240
Db	181	CTTCCTTCCCCCTCCTCGCCGCCACCGCCAGGACC GCCGGCCGGGGGACGAGTCCGGAGC	240
QY	241	agcaggcagaagttaattaaccacttaacctctcagaactgaaacaagaacaacattgtccc	300
Db	241	AGCAGCCAGAGTTATTATAACCACCTTAACCTCTCAGAACTGAACAAGACAACATTTGTCCC	300
QY	301	tggaaagccctctttttaaaaaagtgtagaactttagactlcatagcacgtaataaacctcg	360
Db	301	TGGAAGCCCTCTTTTAAAAAAGGTAGAACCTTAGACTTCATAGCACTGAATTAACCTG	360
QY	361	cactgaaaagctgtttaacctgcatlttgttaactlltgttgaagygaaccaatgtctcaagtt	420
Db	361	CAC TGAAGCTGT TTA C CTGC AT TT GTT G TC ACT TT GTT GA AG TG AC CAT GT CTC AAG TT	420
QY	421	caagltgcaagtlcagaaccocatctgctgctctcctaagygagccaatatcgaacaagaac	480
Db	421	CAAGTGCAAGTTCAGAACCCATCTGCTGCTCTCTCAGGGAGCCAATACTGAACAAGAAC	480
QY	481	cagtcctctctctcacagcccttgaatgatgtattcootctactactagctctctgcccctc	540
Db	481	CAGTCTCTCTCTCACAGCCTTGATGAGATFTCCCTTCTACTACTAGCTCTCTGCCCTCT	540
QY	541	gaaaaatgcaggtagaccocatccaacactctgctcllaaccctctgcatactatlacatccacc	600
Db	541	GAAAATGCAGGTAGACCCATTCAAACACTCTGCTTTACCCCTCTGCATCTATTACATCCACC	600
QY	601	agtcagctgcagaaagcalaacccctactgtagaactaaatgcactylgcatgnaact	660
Db	601	AGTGCAGCTGCAGAAAGCATTAACCCCTACTGTAGAACTAAATGCAC TGTGCATGAAACTT	660
QY	661	ggaaaaaaaccaatgtataagcctgttgacccttactctcgyatgcagtlccacctataac	720
Db	661	GGA AAAAA AACCAATGTATATAAGCCTGT TAGACCTTACTCTCGATGCAGTCCACCTATAAC	720
QY	721	tacaacatgagagagytgtctlatcccccgaggtactttlacccaattccagttccacct	780
Db	721	TACAACATGAGAGAGGTGCTTATCCCCCGAGGTACTTTACCCATTTCAGTTCACACT	780
QY	781	ttactttatcaagtygaaacttctgtgtggaaggacagcaatttaatyygcaaaaggaaagaca	840
Db	781	TTACTTTATCAAGTGAAGCTTTCTGTGTGGAGGACAGCAATTTTAATGGCAAAAGGAAGACA	840
QY	841	agacaggtcgcgaaacacgatgtctgtgcceaaagcgttgaagatccctgcagaatgagccc	900
Db	841	AGACAGGCTGCGAAACACGATGCTGTCTGCCAAAGCGTTGAGGATCCTGCAGATGAGCCC	900

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:45:50 ; Search time 24649.1 Seconds
(without alignments)
727.932 Million cell updates/sec

Title: US-09-316-048-5
Perfect score: 3506
Sequence: 1 acttcctgcggtgctgcggg.....taatactaaaaaaaaaaaaa 3506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_ph:*
- 6: gb_pl1:*
- 7: gb_pl2:*
- 8: gb_pr1:*
- 9: gb_pr2:*
- 10: gb_pr3:*
- 11: em_fun:*
- 12: em_hum1:*
- 13: em_hum2:*
- 14: em_in:*
- 15: em_om:*
- 16: em_or:*
- 17: em_ov:*
- 18: em_pat:*
- 19: em_ph:*
- 20: em_pl:*
- 21: em_ro:*
- 22: em_sts:*
- 23: em_sy:*
- 24: em_un:*
- 25: em_vi:*
- 26: gb_htg1:*
- 27: gb_htg2:*
- 28: gb_in1:*
- 29: gb_in2:*
- 30: em_bal:*
- 31: em_ba2:*
- 32: em_hum3:*
- 33: em_hum4:*
- 34: gb_pr4:*
- 35: gb_htg3:*
- 36: gb_htg4:*
- 37: gb_htg5:*
- 38: gb_htg6:*
- 39: gb_htg7:*
- 40: em_htg1:*
- 41: em_htg2:*
- 42: em_htg3:*
- 43: em_hum5:*

- 44: gb_pl3:*
- 45: gb_pr5:*
- 46: gb_htg8:*
- 47: gb_htg9:*
- 48: gb_htg10:*
- 49: gb_htg11:*
- 50: gb_htg12:*
- 51: gb_htg13:*
- 52: gb_htg14:*
- 53: gb_in3:*
- 54: gb_htg15:*
- 55: gb_htg16:*
- 56: gb_htg17:*
- 57: em_htg4:*
- 58: em_htg5:*
- 59: em_htg6:*
- 60: em_htg7:*
- 61: em_hum6:*
- 62: gb_htg18:*
- 63: gb_htg19:*
- 64: gb_ba3:*
- 65: em_htg8:*
- 66: em_htg9:*
- 67: em_htg10:*
- 68: gb_pr6:*
- 69: gb_pr7:*
- 70: gb_htg20:*
- 71: gb_htg21:*
- 72: gb_htg22:*
- 73: gb_htg23:*
- 74: gb_ro:*
- 75: gb_sts1:*
- 76: gb_sts2:*
- 77: gb_sy:*
- 78: gb_un:*
- 79: gb_vil:*
- 80: gb_vil2:*
- 81: gb_pat1:*
- 82: gb_pat2:*
- 83: em_htg0:*
- 84: gb_htg24:*
- 85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	3502.8	99.9	3506	10	AF061939	AF061939 Homo sapi
2	2958.8	84.4	3349	10	AF061941	AF061941 Homo sapi
3	2914.8	83.1	3217	10	AF061938	AF061938 Homo sapi
4	2896	82.6	3142	10	AF061940	AF061940 Homo sapi
5	2848.8	81.3	3194	34	AK025519	AK025519 Homo sapi
6	2799.4	79.8	3066	68	HSAL132258	AL132258 Homo sapi
7	1357.4	38.7	113836	34	AL133174	AL133174 Human DNA
8	1282.8	36.6	3041	74	AF290989	AF290989 Rattus no
9	1256	35.8	2860	74	AF061942	AF061942 Mus muscu
10	1254.8	35.8	3059	74	AF227200	AF227200 Rattus no
11	705	20.1	157803	35	AC011492	AC011492 Homo sapi
12	553.2	15.8	194407	54	AC068845	AC068845 Homo sapi
13	383.2	10.9	192597	70	AL158850	AL158850 Homo sapi
14	380.2	10.8	385	75	G30939	G30939 DRESS50 Huma
15	291	8.3	160979	52	AC055879	AC055879 Homo sapi
16	287	8.2	360	74	RNO010200	AO1010200 Rattus no
17	283.6	8.1	225016	52	AC063967	AC063967 Mus muscu
18	276.8	7.9	2968	34	AK002152	AK002152 Homo sapi
19	276.8	7.9	4058	68	HSA9062	Y19062 Homo sapien
20	220.2	6.3	263	75	G22107	G22107 human STS W
21	193.6	5.5	163253	68	HSB140E4	AL121877 Human DNA

Db 60919 GTATTTTATTTGGCTGTGTGAAAAAGAAATGGAAGACAGAAAGCAATTAACCTGCTCAC 60860
QY 2237 tcagacacattggygacaaacccctggacagccatgccagagagagagcccttgaccgcccc 2296
Db 60859 ACTGGCTGACTCAAAC-ACCTTGCTATTAACATGCTGGAGAAAAGCCTTAGACTGGCCTC 60801
QY 2297 agagctaaaagcaccagagaaaatcaaatgctcctactcagcgtgacccaactttcta 2356
Db 60800 AAAGCTGAAAAACACCGGGAATAATTAACTGCTTCCTACTCGTTACAACATTAACCTTCTA 60741
QY 2357 gttgccaacgcccccaaccctcctgcagtlacccaaccatcacaccactgtcttctcc 2416
Db 60740 GCATGCTACGG-CCAGTCACCTTACTGCAGTACCCACACCATTAACCACTGCTTCTCT 60682
QY 2417 aacagtgcctgtatctctagtlttcattatcttcttlttgattgatatgacactataaa 2476
Db 60681 AATAGTGATACATATTTTAGTTTATTA-TTCTTTTGATTGAATGATGCCATATGAA 60623
QY 2477 atttcatttggaattctcaattgtatctagttaaatagcacagtttggaaactgtc 2536
Db 60622 ATTTCCATTTGAAAGTTTCTCAGTTGTATCTAGTTACATAGTACAGTTCAGAAACCTGTC 60563
QY 2537 tgagactgacttatacaataatcaaccgacaagaatcatatccatgtgtatgttag 2596
Db 60562 CCAGACTGACTTTATCAGGAATCTAACTGGCATGATCATATCCATGTATACGTGTTGT 60503
QY 2597 acattttattcatatgactaacccagagcagtttcagtgatgcaaatgtgtgccctc 2656
Db 60502 ACATTTCTTTTCTACTGACTAACCAGAACCATTTTCAGTGATGCAAAATCATGTGACCTCT 60443
QY 2657 ggttcagctgaaacagtcctgactt-tcaaaaacctgtaataagtcctccacagttgta 2715
Db 60442 GGTTTAGCTGAAAAAGCCCTGGACTTCTCAGAAACCTTGATCAATTTTCCACTATTGTA 60383
QY 2716 taaattgacaatttaggaattttaactttagatgcatcattgttccattttatttc 2775
Db 60382 CAAATTGGCCATTTAGGAATTTCAAACTCA---ATCATTTGGTCTCTTTTATATTTT 60327
QY 2776 attttattttgttaagca--aacaggacttaaatgaacttgcctcgtttaaag 2833
Db 60326 ATTTTATTTTGTGTTGTTAATGGGACCCTAAACAAATAAATTTGATCTTTTAAAG 60267
QY 2834 attatataaaacacatgtgtatctatatacatatgctcttgagacttagcttcaactaca 2893
Db 60266 ATTATTAATAAACTTACGTGTATGTACATATAGCTCTTCAAGACCCACAGCTTCACTATG 60207
QY 2894 ctacagatatgatctccatgtagtcacatataaaccctgcagagtatttcc-agagtgc 2952
Db 60206 CTGTAGGATATTGTCTCTGTGTAGTACATATGACATATGCAAGTTAATTTTCTTGAAGTGC 60147
QY 2953 tcgatactgttaattacatc 2972
Db 60146 TTTTACTCTTATTTTCTC 60127

RESULT 14
LOCUS G30939 385 bp DNA STS 29-SEP-1998
DEFINITION DRES50 Human DRES Homo sapiens STS genomic clone 22368 3', sequence tagged site.
ACCESSION G30939
VERSION G30939.1 GI:1724032
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 385)
AUTHORS Banfi,S., Borsani,G., Rossi,E., Bernard,L., Gufanti,A.,
Zuffardi,O., Marchitelli,O.A., Giglio,S., Coluccia,E., Zollo,M.,
IDENTIFICATION and mapping human cDNAs homologous to Drosophila

JOURNAL mutant genes through EST database searching
MEDLINE Nat. Genet. 13 (2), 167-174 (1996)
96225443
COMMENT
Contact: Ballabio Andrea
Telethon Institute of Genetics and Medicine, TIGEM
Telethon Institute of Genetics and Medicine, TIGEM
via Olgettina 58, 20132 Milano, Italy
Tel: +39-2-215601
Fax: +39-2-21560220
Email: dres@tigem.it
Primer A: GTAACAATGATGGAAGCAG
Primer B: GGCTGAAAAAGATGACCTAC
STS size: 157
PCR Profile:
Presoak: 94 degrees C for 2 minutes
Denaturation: 94 degrees C for 45 seconds
Annealing: 57 degrees C for 45 seconds
Polymerization: 72 degrees C for 45 seconds
PCR Cycles: 35
Thermal Cycler: MJR PTC-200
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 220 uM
Tag Polymerase: 0.025 units/ul
Total Vol: 15 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
The 5' end of the cDNA clone 22368 is homologous to the staufen Drosophila gene product. EST from which STS is derived: T89105.
Forward primer is T0-1233; Reverse primer is T0-1234.
FEATURES
location/Qualifiers
1..385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="20, 12.10 cR from GCT10F11 (lod >3.0)"
/clone="22368"
/clone_1lb="Human DRES"
29..185
primer_bind 29..49
primer_bind complement(165..185)
BASE COUNT 135 a 79 c 71 g 100 t
ORIGIN
Query Match 11.8%; Score 380.2; DB 75; Length 385;
Best Local Similarity 99.2%; Pred. No. 4.4e-78;
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2780 ttattttgtttaatgcaacagagacttaaatgaacttgcctcgtgtttaaagattat 2839
Db 385 TTTATTTTGTTAATGCAACAGGACTTAATGAACCTTGATCTCTGTTTAAAGATTATT 326
QY 2840 aaaaaacattgtgtatctatatacatatgctcttgagagacttagcttcaactacacag 2899
Db 325 AAAAAACATTGTGTATCTATACATATATGGCTCTTGAAGACCTTAGCCTTCACTACACAG 266
QY 2900 gatagatctccatgtagtcacatataaaccctgcagagtgatttccagagtgctcgatac 2959
Db 265 GATATGATCTCCATGTAGTCCATATATAAACCTGCAGAGTGATTTCCAGAGTGCTCGATAC 206
QY 2960 tgttaattacatctccatagggctgaaaagaatgacctagcttctgtatacagctgtg 3019
Db 205 TGTTAATTACATCTCCATTTAGGGCTGAAAAAGAAATGACCTACGTTTCTGTATACAGCTGTG 146
QY 3020 ttgctttgatgtgtgttactgtacacagaagtggtgacactgagcctcgtgtgtgt 3079
|||||

QY 2111 ggtcgcacagcagctgttctcccggaacaccagcatccacccttgagcactccg 2170
|||||
Db 2946 ggctgccacagcagctggctctctggggcacccagcagatccacgcccctggcacccttcc 2887
QY 2171 tgttgcctttttttttccctgtg--tgaagaagaacgcagcagcccttctcaagc 2228
|||||
Db 2886 tctttggtcttttttttcccggtgaagaagaactgaagcattgaccccttctcaagc 2827
QY 2229 tggctcactcagacacatttggacacaaccccttgacagccatggcagagagagccttga 2288
|||||
Db 2826 tggctta-----gtgggacaaaccttgacacgcccacacagagagagcccttca 2776
QY 2289 ccggcccaagagctaaagcaccagagaaaaatcaaatgcttctactcagcgtgacccaa 2348
|||||
Db 2775 ccagccccagagctaaagcaccagagaaagtcaaatgcttctatttcagcatgacctaa 2716
QY 2349 ctttctagtgccacgagcccaacctctgcagtcaccacacatcaccactgctt 2408
| |||||
Db 2715 c--ttctagtgccacgagcccaacctctgcagtcaccacacatcaccactgctt 2658
QY 2409 tctctcacaagtgatctgtattcttagttcatatttctttgatgatgacac 2468
|||||
Db 2657 tctctccacagctgattctgtattcttagttcatatttctttgatgatgacac 2598
QY 2469 tatataaatctcat 2484
|||||
Db 2597 TATATAAATTTTCTT 2582

RESULT 13
AL158850/c
LOCUS
DEFINITION
Homo sapiens chromosome 6 clone RP3-522B19, ** SEQUENCING IN
PROGRESS **
ACCESSION
AL158850
VERSION
AL158850.4
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 192597)
AUTHORS
Burton, J.
TITLE
Direct Submission
JOURNAL
Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT
On Aug 13, 2000 this sequence version replaced gi:9212635.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dJ522B19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 12% of reads
Sequencing vector: plasmid; L08752; 87% of reads
Chemistry: Dye-terminator Big Dye; 77% of reads
Chemistry: Dye-terminator Big Dye; 22% of reads
Chemistry: Dye-primer-amerham; 22% of reads
Consensus quality: 185625 bases at least Q40
Consensus quality: 188533 bases at least Q30
Consensus quality: 190005 bases at least Q20
Insert size: 191597; sum-of-contigs
Insert size: 144116; 9.6% error; agarose-fp
Quality coverage: 5.71x in Q20 bases; sum-of-contigs Quality
coverage: 8.21x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 13837: contig of 13837 bp in length
* 13838 13937: gap of 100 bp
* 13938 17476: contig of 3539 bp in length
* 17477 17576: gap of 100 bp
* 17577 20005: contig of 2429 bp in length
* 20006 20105: gap of 100 bp
* 20106 32070: contig of 11965 bp in length
* 32071 32170: gap of 100 bp
* 32171 39772: contig of 7602 bp in length
* 39773 39872: gap of 100 bp
* 39873 43471: contig of 3599 bp in length
* 43472 43571: gap of 100 bp
* 43572 51795: contig of 8224 bp in length
* 51796 51895: gap of 100 bp
* 51896 178838: contig of 126943 bp in length
* 178839 178938: gap of 100 bp
* 178939 181919: contig of 2981 bp in length
* 181920 182019: gap of 100 bp
* 182020 184231: contig of 2212 bp in length
* 184232 184331: gap of 100 bp
* 184332 192597: contig of 8266 bp in length.
FEATURES
source
1. 192597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-522B19"
/clone_lib="RPCT-3"
1. 13837
/note="assembly_fragment:02383
fragment_chain:1"
13938. 17476
/note="assembly_fragment:01080
fragment_chain:1"
17577. 20005
/note="assembly_fragment:02434
fragment_chain:1"
20106. 32070
/note="assembly_fragment:03060
fragment_chain:2"
32171. 39772
/note="assembly_fragment:00945
fragment_chain:2"
39873. 43471
/note="assembly_fragment:00013
fragment_chain:2"
43572. 51795
/note="assembly_fragment:02225
fragment_chain:3"
51896. 178838
/note="assembly_fragment:01767
fragment_chain:3"
178939. 181919
/note="assembly_fragment:01637"
182020. 184231
/note="assembly_fragment:02408"
184332. 192597
/note="assembly_fragment:02960"
BASE COUNT 58324 a 39515 c 38313 g 55431 t 1014 others
ORIGIN

QY 2177 gctcttttttccctgtgtgaagaagaacgcagcagcccttctcaagctgctcac 2236
Query Match 11.9%; Score 383.2; DB 70; Length 192597;
Best Local Similarity 73.2%; Pred No. 1.7e-78;
Matches 586; Conservative 0; Mismatches 203; Indels 11; Gaps 7;

*	1856	3429:	contig of 1574 bp	in length
*	3430	3529:	gap of 100 bp	
*	3530	5346:	contig of 1817 bp	in length
*	5347	5446:	gap of 100 bp	
*	5447	7338:	contig of 1892 bp	in length
*	7339	7438:	gap of 100 bp	
*	7439	10297:	contig of 2859 bp	in length
*	10298	10397:	gap of 100 bp	
*	10398	13590:	contig of 3193 bp	in length
*	13591	13690:	gap of 100 bp	
*	13691	16293:	contig of 2603 bp	in length
*	16394	16393:	gap of 100 bp	
*	16394	19065:	contig of 2672 bp	in length
*	19066	19165:	gap of 100 bp	
*	19166	22713:	contig of 3548 bp	in length
*	22714	22813:	gap of 100 bp	
*	22814	27165:	contig of 4352 bp	in length
*	27166	27265:	gap of 100 bp	
*	27266	32377:	contig of 5112 bp	in length
*	32378	32477:	gap of 100 bp	
*	32478	36572:	contig of 4095 bp	in length
*	36573	36672:	gap of 100 bp	
*	36673	41657:	contig of 4985 bp	in length
*	41658	41757:	gap of 100 bp	
*	41758	47024:	contig of 5267 bp	in length
*	47025	47124:	gap of 100 bp	
*	47125	52938:	contig of 5814 bp	in length
*	52939	53038:	gap of 100 bp	
*	53039	59129:	contig of 6091 bp	in length
*	59230	59229:	gap of 100 bp	
*	59230	65896:	contig of 6667 bp	in length
*	65897	65996:	gap of 100 bp	
*	65997	89427:	contig of 23431 bp	in length
*	89428	89527:	gap of 100 bp	
*	89528	95785:	contig of 6258 bp	in length
*	95786	95785:	gap of 100 bp	
*	95886	102762:	contig of 6877 bp	in length
*	102763	102862:	gap of 100 bp	
*	102863	110899:	contig of 8037 bp	in length
*	110900	110999:	gap of 100 bp	
*	111000	121002:	contig of 10003 bp	in length
*	121003	121102:	gap of 100 bp	
*	121103	133048:	contig of 11946 bp	in length
*	133049	133148:	gap of 100 bp	
*	133149	145755:	contig of 12607 bp	in length
*	145756	145855:	gap of 100 bp	
*	145856	165006:	contig of 19151 bp	in length
*	165007	165106:	gap of 100 bp	
*	165107	188243:	contig of 21337 bp	in length
*	188244	188343:	gap of 100 bp	
*	188344	194407:	contig of 6064 bp	in length.

FEATURES
source

```
misc_feature 1. .1755 /note="assembly_fragment"  
misc_feature 1856. .3429 /note="assembly_fragment"  
misc_feature 3530. .5346 /note="assembly_fragment"  
misc_feature 5447. .7338 /note="assembly_fragment"  
misc_feature 7439. .10297 /note="assembly_fragment"  
misc_feature 10398. .13590 /note="assembly_fragment"  
misc_feature 13691. .16293 /note="assembly_fragment"
```

misc_feature	16394. .19065
	/note="assembly_fragment"
misc_feature	19166. .22713
	/note="assembly_fragment"
misc_feature	22814. .27165
	/note="assembly_fragment"
misc_feature	27266. .32377
	/note="assembly_fragment"
misc_feature	32478. .36572
	/note="assembly_fragment"
misc_feature	36673. .41657
	/note="assembly_fragment"
misc_feature	41758. .47024
	/note="assembly_fragment"
misc_feature	47125. .52938
	/note="assembly_fragment"
misc_feature	53039. .59129
	/note="assembly_fragment"
misc_feature	59230. .65896
	/note="assembly_fragment"
misc_feature	65997. .89427
	/note="assembly_fragment"
misc_feature	89528. .95785
	/note="assembly_fragment"
misc_feature	95886. .102762
	/note="assembly_fragment"
misc_feature	102863. .110899
	/note="assembly_fragment"
misc_feature	111000. .121002
	/note="assembly_fragment"
misc_feature	121103. .133048
	/note="assembly_fragment"
misc_feature	133149. .145755
	/note="assembly_fragment"
misc_feature	145856. .165006
	/note="assembly_fragment"
misc_feature	165107. .188243
	/note="assembly_fragment"
misc_feature	188344. .194407

	vector_side:right"									
BASE COUNT	50993	a	43965	c	46383	g	50458	t	2608	others
ORIGIN										

Query Match	17.2%;	Score 553.2;	DB 54;	Length 194407;
Best Local Similarity	90.4%;	Pred. No. 3.9e-118;		
Matches 665; Conservative	0;	Mismatches 48;	Indels 23;	Gaps 6;

OY	1752	gctgcgctgaacatcctaagttgtctctgtagtgtgaccaacaagaatacagatatgccaa	1811
Db	3297	GCTGCACCTGAACACACTTAAGTTCCTGTCTGAGTTGGACCACAAGAATAACAGAGATGCCA	3238
OY	1812	agaacaaggaaaacgagccaattgtctgtgtgtgtggagtgctgaaccttctcgtgcatgaa	1871
Db	3237	AGAACAGGAATGGATCAATGTCTGTGTGTGGAGGTGCTGAACCTTTCTCGCCATGAA	3178
OY	1872	ccatataaatacccacacatatataactgaanaatactgaanaactgctttaaaaaatttgyaat	1931
Db	3177	CCAATTATAAAATCCCAACAFATACACTGAAAAATATTAAAACCTGCTTGAAAAATTGGAAAT	3118
OY	1932	tctcgatatcctecagttggtgcgcgagagacacagtytggttaangatgttggcagcacgaggt	1991
Db	3117	TTCCTGATACCTCTAGTGGGCCGAGAG--ACGGTGGGTAAGAAGATGTGGGCAGCAGCAGGG	3060
OY	1992	aagacaacagaaaacacacaagagtcggtctgtggcc-ggctgtgactgtgtcgtgggtttgttg	2050
Db	3059	AACACACAACAG-----GGGAGGTGGCTGTGGCCCGGCGCTGGACTGTGCTGGGGTTTGTTG	3007
OY	2051	tgatgtgccactcgtgtgacctgtgctgtgtccctaacgcaatalagcagctgtcctgttgggaaagaag	2110
Db	3006	TGACGGCCACTCGGTGACCTGGCTGTTCCTACGCAATACCAAGCTGCTTTGGGGAAAGAAG	2947

Db 150307 C--TTCTAGTGTGCCACAGGCCCCACCACTTCTCGAGTGGCCACACCATCACCACCTGCTT 150250
QY 2409 tctctccaacagtgtatctgtatctttagtlttcattatcttcttgaatgatat----- 2463
Db 150249 TCTCTTCCAAAGTGTATCTTCTAGTTTCATTATTTCTTTTATTGGTATGACAC 150190
QY 2464 ----- 2463
Db 150189 TATATAAAATTTTCCCTCCCTGCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 150130
QY 2464 ----- 2463
Db 150129 TTCTTTCTTCAAGACAAGTCTCAGCTCTGTCAACCAGGCTGGAGTGCACGCGACAGTC 150070
QY 2464 ----- 2463
Db 150069 TTGGCTCCCTGCAACTTCTGCTCCAGGTTCAAGCAATTGCTGCTCAGCCTCCGAG 150010
QY 2464 ----- 2463
Db 150009 TAGCTGGGCCACACACCCAGCTAGTTTGTGACTTTTAATAGAGACGGGGTTTCAACAT 149950
QY 2464 ----- 2463
Db 149949 CTTGGCCAGGCTGGTCTTGAACCTCTGACCTCGTGATCCACCCGCTCGGCTCCCAAG 149890
QY 2464 ----- 2463
Db 149889 TGCTGGGATTACTGGCATGAGCCACCGTGCCCGACGACACTATATAAATTTTCATTGGA 149830
QY 2489 gaatttctcaattgtatctagttaaatagcacagtttgaaacttgtctgagactgactt 2548
Db 149829 GAGTTTCTCAACTGTATCTAATTAATCGCACAGTTTGGAAACTTTTCTGAGACTGACTT 149770
QY 2549 tatcaataatctaacccgacaagaatcatatccatgtgtatgtgttagacatttattt 2608
Db 149769 TATCAATTAATCTAACCCGACAAGATCATATCCAGCTGTATGTAGACATTTTATTTT 149710
QY 2609 catgtactaacccaggaacagttcagtgtatgcaaatgtgtgccccctgtgtcagctgaa 2668
Db 149709 CATTGACTAACCCAGCACCATTTCAAGTATGACAAATTTAGCTGCCCCCTATGTGACGTGAA 149650
QY 2669 acagtccctgacttccaacaaacttgaataagtctccacagttgtataaatlgacaat 2728
Db 149649 ACAGTCTCTGCACTTTCAAAAACCTTGAATTAAGTCTCCACAGTTGTATAAATTGACAAT 149590
QY 2729 ttaggaatttaaaacttagatgatcatattgttccattttatttatttattt 2788
Db 149589 TTAGGAATTTTAAACTTTAGATGATCATTTGCTTCCATTTTATTTTATTTTATTTTG 149530
QY 2789 ttaatgcaaacagygact----aaatgaactttgatctctgtltttaaagattatataaaa 2844
Db 149529 TTAATGCAACACGAGACTTAAACAATGAACCTTGATCTCCATTTTAAATATTCTTTAAGA 149470
QY 2845 acatgtgtatctatacatatgctcttgaggacttagcttccactacactacagatat 2904
Db 149469 AAATCATGTGTCTATACATACGGCTCTTGAGGAGCTTAGCTTTCACCTACACTACAGATATP 149410
QY 2905 gatctccatgtagtcacataataaactgcagagtgtatttccagagtgctcgatactgtta 2964
Db 149409 GATCTCCACATAGTTACATATAAAGTGCAGACTGATTTTCCAGAGTGTGCTGATACATATTA 149350
QY 2965 attacatcccatctagtgctgaagaagaatgacactacgttctgtatatacagctgtgtgct 3024
Db 149349 ATTACATCTCGATTAGGGCTGAAGAAGTATGACCTATGTTCTTTATACAGTTGTGTGCT 149290
QY 3025 ttgatgtgtgttactgtacacagaagtgtgtgcactgaagc 3067
Db 149289 TTTGATGTTGTGTACTATACAGAGAAGTGTGTGACACTGGGCC 149247

RESULT 12
AC068845/c

LOCUS AC068845 194407 bp DNA HTG 04-SEP-2000
DEFINITION Homo sapiens chromosome 19 clone RP11-73711 map 19, WORKING DRAFT
SEQUENCE, 27 unordered pieces.
AC068845
VERSION AC068845.3 GI:9966312
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 194407)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 19, clone RP11-73711
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194407)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trifoglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION

TITLE Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 4, 2000 this sequence version replaced gi:8247880.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7549
Center clone name: 737_I_1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181875 bases at least Q40
Consensus quality: 187842 bases at least Q30
Consensus quality: 190220 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 191807; sum-of-ctnigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1755: ctnig of 1755 bp in length
* 1756-1855: gap of 100 bp

RESULT 10
LOCUS AF227200 3059 bp mRNA ROD 09-AUG-2000
DEFINITION Rattus norvegicus staufen isoform Stau+I6 (Stau) mRNA, complete cds, alternatively spliced.
ACCESSION AF227200
VERSION AF227200.1 GI:9754865
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3059)
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desgroseillers, L., Kuhl, D., Richter, D. and Kindler, S.
TITLE Two rat brain staufen isoforms differentially bind RNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3059)
AUTHORS Monshausen, M. and Kindler, S.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) University of Hamburg, Institute for Cell Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg 20246, Germany
FEATURES
source Location/Qualifiers
1..3059
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1..3059
/gene="Stau"
227..1714
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau+I6"
/codon_start=1
/product="staufen isoform Stau+I6"
/protein_id="AAF98119.1"
/db_xref="GI:9754866"
/translation="MYKVPDPSHRMSTSYSGMRGAYPPRYFPVPVPLLYQVELS VGGQFNGKMRPVPKHDATARALRTLOSEPLPERLEVNRESEENLNKSEISOVF EIALKRNLPVNESEFPLTOVARESGPRHMKNFVTRVSVGEFGESEKSKISKNNRA RAVLEQLRLPLPAVERKPRIKKSQPTCKIQTAPDYOGMNPISRLAQIOQAKKE KEPEYMLTERGLPRRREFVMQVKGHHTAEGAGTNKKVAKNNAENMLEILGFKVPQ AOPAKPALKSEKTPVKPGDGRKVTFFEPSPGDENGTSNKEDFERMPYLSHOOLPAG ILPMVPEVAQAVGVSGHHTKDFARAPNPAAKATVTAMIARELLYGTSPTAETILKS NISSGHVPHGPRTPEQLYLSRAQFQVEYKDFPKNNKNECVSLINCSSQPLVSH GIGKDVESHDMALNILKLSLSDQSTEMPRTGNGPVSACGRC"
polyA_signal 3043..3048
BASE COUNT 817 a 778 c 749 g 713 t 2 others
ORIGIN
Query Match 39.0%; Score 1253.8; DB 74; Length 3059;
Best Local Similarity 86.3%; Pred. No. 9.7e-282;
Matches 1428; Conservative 0; Mismatches 207; Indels 20; Gaps 3;

Db 346 AGTTAGACTCTCCGTGGAGAGACAGCTTTAATGGAGGAAAGATGAGACCACTGT 405
QY 563 gaacacgatactgctgcgaagcgttgagatccctgcagaatgagccctgcagagag 622
Db 406 GAAACATGATGCCACTGCCCGAGCCCTGAGACTCTGCAGAGTGAAGCCCTACAGAGAG 465
QY 623 gctgaggtgtaatggaagaaatccgaagaagaataatcctaataatctgaataagtca 682
Db 466 GCTAGAGGTAAATGGAAGAGATCAGAAGAAGAAACCTTAATAATCAGAATAAGCCA 525
QY 683 agtgttgagattgcacttaacggaacttgccctgtgaattcg----- 726
Db 526 AGTGTTTGAGATTGCCGTGAAGCGGAATTTGCCCTGTGAATTTTGAGCTTTCCCTCTGAC 585
QY 727 --agtgcccgaggagagtgccccacacatgaagaacttltgacaaagtttcggt 784
Db 586 ACAGGTGGCCCGGAGAGTGGCCGCCACACATGAAGAATTTGTGACCAGGGTTTCAGT 645
QY 785 tgggagattgttgggggaagtggaaggaaagcaagaagatttcaagaanaaatgcgcg 844
Db 646 TGGGAGTTTGTAGGGGAAGAGAGGAAAGCAAGAGATCTCCAAGAAGAAATGCGGC 705
QY 845 catagctgttcttgagagagctgaagaagttaccgccccctgcctgcagtttgaacgagtaa 904
Db 706 AAGGCCCTGCTGGAGCAGCTAAGAGGCTGCCACCCTTACCTGCTGTGGAGCGAGTGAA 765
QY 905 gccatgaatcaaaaaaagaaacaaacccatagtcgaagccacagacaagccagataatgg 964
Db 766 GCCCAGAAATCAAGAAAGAAAGTCAAGCCCACTGCAAGCTTACAGACTGCCCCCGACTATGG 825
QY 965 ccagggatcaatccgattagccgactgcccagatccagcagggcaaaaaaggaagaa 1024
Db 826 CCAAGGATGAATCCTATTAGCAGACTGGCACAGATCCACAGGCAAGCAAAAAAGCAAGAA 885
QY 1025 gccagagtaacgctctcacaagacgagcgctcccgccgagggaggttggatgca 1084
Db 886 GCCGAGTACATGCTCTTACAGACGAGGTCTTCCACGTCGCAAGGAGTTGTATGCA 945
QY 1085 ggtgaaggttggaacaacacactgcagaaggaacggcgacccaacaagaagttgccaagcg 1144
Db 946 GGTGAAGGTGGGCATCACACCCGACAGAAGGACAGGTACCAATAAGAAGGTGGCCAAAGCG 1005
QY 1145 caatgcagccgagaacatgctgtagatccctgttccaagtcccgacggcgagccac 1204
Db 1006 CAATGCTGCTGAGAACATGCTGAGATCCTGGGGTTCAAAAGTCCCCAGGCGCAGCCTGC 1065
QY 1205 caaacccgactcaagtacagagagaagacaaccataagaaaacaggggatggaagaaa 1264
Db 1066 CAAGCCAGCACTCAAAATCAGAAGAGAAGACTCCAGTAAGAAGAACAGAGATGAAGGAA 1125
QY 1265 agtaaccttttgaacctgctcgtggagatgaanaatgagactagtaataaagagatga 1324
Db 1126 AGTAACATTTTGAACCTAGCCCTGGGGATGAAAAATGGAACCAATAAAGAGATGA 1185
QY 1325 gttcagatgcttatcctaagtaatcagcagctgctgcgtggaattcttcccatgtgcc 1384
Db 1186 GTTCAGAGATGCCCTATCTTAGCCATCAGCAGCTGCCAGCGGGATTTCTCCCAATGGTGCC 1245
QY 1385 cgaagtcgcccagagctgttagagtttagtcaagagacatcacaccaagaatttaccagggc 1444
Db 1246 AGAGGTTGCCCAGGCTGTGGGGTTAGTCAAGGACACCATACCAAGATTTTCGCCAGGGC 1305
QY 1445 agctccgaatcctgcgaagggccaggtlaactgccatgatagcccgagagttgtgtatgg 1504
Db 1306 AGCTCAAAATCTCTGCCAAGGCCACGGTAACCTGCCATGATAGCCCGAGAATTTGTGACGG 1365
QY 1505 gggcaccctgcgccacagccgagacatttlaaagataacatcttcagggcaggtacc 1564
Db 1366 GGGCACCCTGCCACAGCCGAGACCAATTTTAAAGAGTAACATCTCTTCAGGACATGTACC 1425
QY 1565 ccatgacctctcacagagacctctgacaaactgtaactgtatcttccagagtcaggagatt 1624
|||||

Db 166 AGAAGTATAGCTTCTACCATTTGAGCTCAATGCACTGTGCGTGAACCTGGAAGAAAACC 225
QY 383 aatgtataagcctgttgacccttactctcgatgcagtcaccctataactacaacatgag 442
Db 226 AATGTATAAGCCCGTGGACCCCCACTCTCGGATGCGATCCACCCTACAGCTATGGCATGCG 285
QY 443 aggaagtgctatcccccgaggtactttlaccattccagtlccacccttacttatca 502
Db 286 TGGAGGTGCTTATCCCCAGATACTTTTACCATTTCAGTTCCACCGTTACTCTACCA 345
QY 503 agtgaacttctgttgaggagacagcaatttaatgycaaagaaagacagacagctgc 562
Db 346 AGTTGAGCTCTCCGTGGAGGACAGCAGTTTAATGGGAAGGAAAGATGAGACCACCTGT 405
QY 563 gaaacacgatgtctgtgccaaagcgttgagatccctgcagaatagcccctgcagagag 622
Db 406 GAAACATGATGCCACTGCCCCGAGCCCTGAGGACTCTGCAGAGTGAGCCCTTACCAGAGAG 465
QY 623 gctggaggtgaatggaagagaatccgaaagaaaaaactcaataaatctgaataagtca 682
Db 466 GCTAGAGGTAATGGAAGAGATCAGAAGAGAAAACCTTAATAATCAGAAATAAGCCA 525
QY 683 agtgtltgagatlgcacttaaacggaactgtgctgtlaattlcgagtgycgcgagag 742
Db 526 AGTGTTTGAGATTGCGTGAAGCGGAATTGCTGTGAATTTTGAGGTGGCCCGGAGAG 585
QY 743 tggccacccccacatgaagaactttgtgaccaaggtttcgtgttgaggagttgttgagg 802
Db 586 TGGCCCGCCACACATGAAGAACCTTTGTGACCAGGGTTTCAGTTGGGAGTTGTAGGGA 645
QY 803 aggtgaagggaagaagaagaatttccaagaaaaatgcccacatagctgttcttgaga 862
Db 646 AGGAGAAGGGAAGAAAGCAAGAAGATCTCCAAGAGAATGCGGCAAGGCGCTGCTGGAGCA 705
QY 863 gctgaagaagtltaccgccccctgctgtcagtlgaacgagtlaaagcctagaatcaaaaaga 922
Db 706 GCTAAGGAGGCTGCCACTCTTACTGCTGTGGAGCGAGTGAAAGCCCAAGATCAAGAAGAA 765
QY 923 aacaaaaccatagtcgaagccaagacaagaagcccaagaatatgcccagggatcaatccgat 982
Db 766 AAGTCAGCCCACTGCAGACTTACAGACTGCCCCGCACTATGGCCAAGGATGAATCCTAT 825
QY 983 tagccgactgcccagatccagcagcgcaaaaaaggaagagccagagtlacacgctcct 1042
Db 826 TAGCAGACTGGCACAGATCCAGCAGGCAAAAAGGAAGAGCCGGAGTACATGCTCCT 885
QY 1043 cacagagcagggcctcccgcgccgcagggagtttgtatgcaggtgaaggttgaaacca 1102
Db 886 TACAGACGAGGCTTTCACAGCTCGCAGGGAGTTTGTGATGCAGGTGAAGGTTGGGCATCA 945
QY 1103 cactgcagaagaagcggcaccacaagaaggtggccaagcgcaatgcagccgagaaacat 1162
Db 946 CACCGCAGAAAGAGAGGTACCAATTAAGAAGTGGCCAAAGCGCAATGCTGTGAGAACAT 1005
QY 1163 gctggagatccttggtttcaaatgccgcagcgcgagcccaaccaaccgcgactcaagtc 1222
Db 1006 GCTGGAGATCCTGGGTTCAAGTCCCGCCAGCGGCAGCCTGCCAAGCCAGCACTCAATC 1065
QY 1223 agagggagaagacaccataaagaagaaccaggggatgtgaagaaaaagtaacctttttgaacc 1282
Db 1066 AGAAGAGAGAGACTCCAGTAAAGAAACCAGGAGATGGAAGAAAGTAAACATTTTGAACC 1125
QY 1283 tggctctgggataaaaaatgggactagtaataaagagatgagtlcagatgccttatct 1342
Db 1126 TAGCCCTGGGATGAATAATGGAACCAAGTAAAGAGATGAGTTCAGGATGCCCCATATCT 1185
QY 1343 aagtcatacagcagctgcctgtgtaattcttcccatggtgcccgaggtcgccagagctgt 1402
Db 1186 TAGCCATCAGCAGCTGCCACCGGGATTCCTCCCATGTTGCCAGAGGTTGCCAGGCTGT 1245
QY 1403 aggaagtagtcaaggacatcacaccaagaatttaccagggcagctccgaatcctgcacaa 1462
Db 1246 CGGGGTTAGTCAAGGACACCATACCAAGAATTTTCGCCAGGGCAGCTCCAATTCCTGCCAA 1305

QY 1463 ggcacacggttaactgcatgatagcccgagagtggtgtatggygggaccccgccacagc 1522
Db 1306 GGCCACGGTAACTGCCATGATAGCCCGAGAATTTGTTAGCGGGGCACCTCGCCACAGC 1365
QY 1523 cgagaccattttaagaataacatctcttcagcgccacgtaaccatgtaacctctcacgag 1582
Db 1366 CGAGACCATTTTAAAGAGTAACTCTCTTCAGGACATGTACCCCATGAGCTCGCACAG 1425
QY 1583 accctctgagcaactgactatcttccagaggtccaaggatccaggttgaataacaaga 1642
Db 1426 ACCCTCTGAGCAACTGTACTACTTTCAGAGAGCCCGAGGATTCAGGTTGAATACAAAGA 1485
QY 1643 ctcccccaaaaaacaagaacgaalttgtlatcttatacaatgtctcctctcagccacc 1702
Db 1486 CTTCGCCAAGAACAAACAGAGAGTGTATCTCTCATCAACTGCTTTCACAGCCACC 1545
QY 1703 tctgatcagccatggtatcggcaaggaatgtgagtcctgcacatgatatggtcgtgaa 1762
Db 1546 TCTCGTCAAGCCATGCGATCGGCAAGGATGTGAGTCCCTGTCATGATATGGCTGCACCTGAA 1605
QY 1763 catcttaaagttgtctgtgttgatgtgaccacaagaatcacagatgccaagaacagaaa 1822
Db 1606 CATCTTAAAGCTGCTGTCTGAGTTGGACCAACAGACACAGAGATGCCAAGACAGAAA 1665
QY 1823 cggacccaatgtctgtgtggygaggtgtgtgaacctttctgycatgaaccattataaa 1882
Db 1666 CGGACCAAGTTTACGCGTGGCGGAGGTGCTGAACCTATTCTGGCCATGAACATTATAAAA 1725
QY 1883 tcccaacatatatactgaaaatactga-aactgctttgaaaaatttgaattctgtatacc 1941
Db 1726 -CCCAACATATATACCGAAATACTGAGAACCTGTTGAAAAATTGGAAATTCTGATAAC 1784
QY 1942 tccagtggygccgagaga 1958
Db 1785 TCCAGTGGGCCAAGACA 1801

RESULT 9
AF061942 LOCUS 2860 bp mRNA ROD 04-MAR-1999
DEFINITION Mus musculus RNA-binding protein stauflen mRNA, complete cds.
ACCESSION AF061942
VERSION AF061942.1 GI:4335944
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2860)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.
TITLE Mammalian stauflen is a double-stranded-RNA- and tubulin-binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 2860)
AUTHORS Duchaine,T., Luo,M. and DesGroselliers,L.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES
source location/Qualifiers
1. .2860
/organism="Mus musculus"
/db_xref="taxon:10090"
326. .1789
/codon_start=1
/product="RNA-binding protein stauflen"
/protein_id="AADI7529.1"
/db_xref="GI:4335945"
/translation="MYKPYVDPSHRMOSTYSYGMRGAYPPRYFPVPVPLLYQVELS
VGGQFNKGKMRPVKHDAFALRTLQSEPLPERLEVNGREAEENLNKSEISQVF
EIALKRNLPVNEFVARESGLPHKNFVTRVSGEFGSEGSKSKISKKNAAAVLEQ
LRRLLPDAVERVKPRIKKKSQPTCKTAPDYGGMNPISRLLAQIOAKKKEKPEYMLL

Db 34845 ACACGGTGGTAAGGATGTGGGACAGCAGGGAAGACACAGAAACACAGAGGCGG 34786
Qy 2018 ctgtggcc-ggctggactgtgtggtgtgtgtatggtccactcgtgacctggcggt 2076
Db 34785 CTGTGGCCGGGCTGGACTGTGCTGGGGTTTGTGTGATGGCCACTCGGTGACCTGGCGGT 34726
Qy 2077 ccctacgcaatagcagctgctgtggtggaagaagggtgccccagcagctggtctccg 2136
Db 34725 CCCTACGCAATAGCAGCTGCCTGTGGGGAAGAAGGCGTGCACAGCAGCTGGTCTCCG 34666
Qy 2137 ggacaccagcagatccacacccctgggcacctccgtgtgtgtctttttccctgtgt 2196
Db 34665 GGACACCAGCAGATCCACACCCCTGGGACCTCCGTGTTGCTTTTTCCTCCCTGTGT 34606
Qy 2197 gaaagaagaacgcaacgcaacccctctcaagctggtcactcagacacattgggacaac 2256
Db 34605 GAAAGAAGAACGGCACGACCCCTTCTCAAGCTGGCTCACTCAGACACATTGGGACAAAC 34546
Qy 2257 cctggacagccatgccaagagagaggcctttgacccggccccagagctaaagaaccagaga 2316
Db 34545 CCTGGACAGCCCATGCCAGAGAGAGAGGCCCTTTGACCGGCCCCAGAGCTAAAGCACCAGAGA 34486
Qy 2317 aaatcaaatgcttctactctcagcgtgacccaacttctcagtgtgccacggccccaccac 2376
Db 34485 AAATCAAAATGCTTCTACTCTCAGCGTGACCAACTTTTCTAGTGTGCCACGGCCCCACACAC 34426
Qy 2377 ctctcgagtaaccacacacacacacacactgtctctcctcacaagtgatctgtattcta 2436
Db 34425 CTCTCGAGTACCCACACACCATCACACACTGCTTCTCTCCAAACAGTATCTGTATTCTTA 34366
Qy 2437 gttcattatttctctttgattgatgatgacacataaaatttcaatttgagaatttct 2496
Db 34365 GTTTCATTATTCTTTGATTGATATGACACTATATAAAATTTTCATTGAGAATTCT 34306
Qy 2497 caattgtatctagttaaatagcacagtttgaaaacttgtctgagactgactttatcata 2556
Db 34305 CAATTGTATCTAGTTAAATAGCACAGTTTGGAAACTTGTCTGAGACTGACTTTATCAATA 34246
Qy 2557 atctaaccgacaagaatcatalatcatgtgtatgtgttagacattttatttcatgtact 2616
Db 34245 ATCTAACCGACAAGATCATATTCATGTGTGTGTAGACATTTTTATTTCATTGACT 34186
Qy 2617 aaccacagacagtttcaagtgalgcaaatgtgtgccctcgtgtcagctgaaacagtcct 2676
Db 34185 AACCCAGGACAGTTTCAAGTGAAGTCTCCCAAGTTGTATAAATGGACAATTAGGAAT 34126
Qy 2677 ggaacttcaaaaaccttgaataagctccacacagttgtataaatgtgacaatttgaat 2736
Db 34125 GGACTTTCAAAAACCTTGAATAGTCTCCCAAGTTGTATAAATGGACAATTAGGAAT 34066
Qy 2737 tttaaacttagatgatcatgttgttccatttttatttatttatttatttatttatttatt 2796
Db 34065 TTTAACTTTAGATGATCATTTGGTTCATTTTATTTCATTTTATTGTTAATGCA 34006
Qy 2797 aacaggaacttaaatgaactttgatctcgttctttaaagattatataaaaaaacattgtgtatc 2856
Db 34005 AACAGGACTTAATGAACTTTGATCTCTGTGTTTAAAGATTATTAAAAACATTGTGTATC 33946
Qy 2857 tatacatatgctctgtgaagacttagcttctcactacactacagatatgatactcctcatgta 2916
Db 33945 TATACATATAGGCTCTTGAGGACTTAGCTTTTCACTACACTACAGGATATGATCTCCATGTA 33886
Qy 2917 gtccatataaacctgacagtgatlttccagagtgctcgatactgttaattacatctcca 2976
Db 33885 GTCCATATAAACCTGCAGAGTGATTTTCCAGAGTGCTCGATACTGTTAATACATCTCCA 33826
Qy 2977 ttagggctgaaaagaatgacctacgttctgtatcacagctgtgtgtcttgatgtgtg 3036
Db 33825 TTAGGGCTGAAAAAAGAAATGACCTACGTTTCTGTATACAGCTGTGTGTTGATGTTGTG 33766
Qy 3037 ttaactgtacacagaagtgtgtcactgaggtctgtcggtgtgtgtccgtatgtaaaacctg 3096

Db 33765 TTACTGTACACAGAAGTGTGTGCACCTGAGGCTCTGCGTGTGTCGGTATGGAAAGCCTGG 33706
Qy 3097 tagccctgcagtaagtaactgtcttccatcattgttlaagctggaatttctcccat 3156
Db 33705 TAGCCCTGCAGATTAAAGTACTGCTTCATTCATTGTTTACGCTGGAATTTTCTCCCAT 33646
Qy 3157 ggaatgtaagtaaaacttaagtglttgcatacaataatgtaactactaa 3207
Db 33645 GGAATGTAAGTAAACTTAAGTGTGTCATCAATAATGTAATACTAAA 33595
RESULT 8
AF290989 3041 bp mRNA ROD 09-AUG-2000
LOCUS Rattus norvegicus stauven isoform Stau-I6 (Stau) mRNA, complete
DEFINITION cds, alternatively spliced.
ACCESSION AF290989
VERSION AF290989.1 GI:9755162
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE 1 (bases 1 to 3041)
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M.,
DesGroseillers, L., Kuhl, D., Richter, D. and Kindler, S.
TITLE Two rat brain stauven isoforms differentially bind RNA
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 3041)
TITLE Monshausen, M. and Kindler, S.
JOURNAL Direct Submission
FEATURES
source Submitted (28-JUL-2000) Institute for Cell Biochemistry and
Clinical Neurobiology, University of Hamburg, University Hospital
Eppendorf, Sudefeldstr. 24, Hamburg D-22529, Germany
1. 3041
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1. 3041
/gene="Stau"
227. 1696
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila stauven;
contains four RNA-binding domains; rStau-I6"
/codon_start=1
/product="stauven isoform Stau-I6"
/protein_id="AAF98177.1"
/db_xref="GI:9755163"
/translation="MYKPVDSHRMSTYSYGMRGAYPPRYFPVPPVLLYOVELS
VGQGFNGKGMRPVKHDAVARALRTLOSEPLPERLVNGRESEENLNKSEISOVF
EIALKRLPVPNFEVARESGPPHKNFVTRVSVGEFVGESEKSKISKNAARAVLEQ
LRLRLPDAVERVKPRIKKKSQPTCKLQTAPDYGGMNPISRLAQIOAKKEPEYM
LITERGLPRRREFVMQVKGHTAEGAGTNKKVAKRNAEAMLEILGFYPOAOPAKP
ALKSEKTPVKKPGDKRVTFEPSPGDENGTSNKEDEFMRPYLSHQQLPAGILPMVP
EVAQAVGVSOGHHTKDFARAAPNPAKATVTAMIARELLYGGTSPYAEILLKSNISSGH
VPHGPRTRPSQOLYLSRAQGFQVEYKDFPKNNKNECVSLINCSQPLVSHGIGKDV
ESCHDMAALNLIKLSLSDQSTEMPTNGPVSACGRG"
3025. 3030
polya_signal
/gene="Stau"
BASE COUNT 814 a 771 c 747 g 707 t 2 others
ORIGIN
Query Match 39.8%; Score 1281.8; DB 74; Length 3041;
Best Local Similarity 87.28; Pred. No. 2.9e-288;
Matches 1428; Conservative 0; Mismatches 207; Indels 2; Gaps 2;

Db 1919 ATGGGACAAA-CCTGGACAGCCATGCCAGAGAGAGG-C TTTGACCGGCCCCAGAGCTAA 1976

Qy 2305 aagcaccagagaaaaatcaaatgcttcctactcagcgtgacccaacttctctagtgtcca 2364

Db 1977 AAGCACGAGAGAAAATCAAAATGCTTCTTACTCAGCGTGACCCAACTTTTCTAGTGTGCCA 2036

Qy 2365 cggccccaccacctcctgcagtaaccacacacatcacacgtcttctctccaacagtga 2424

Db 2037 CGGCCCCACCACCTCCTGCAGTACCACACCATCACCACTGCTTCTCTCCACAGTGA 2096

Qy 2425 tctgtatctcttagtttcattatcttcttgattgatgtgacactatataaaatttcat 2484

Db 2097 TCTGTATTCTTAGTTTCATTAATTTCTTTGATTGATATGACACTATATAAAATTTTCAT 2156

Qy 2485 ttgagaatttctcaattgtatctagttaataatagcacagtttgaaactgtctgagactg 2544

Db 2157 TTGAGATTTCTCAATGTATCTAGTTAAATAGCACAGTTTGGAACCTGTCTGAGACTG 2216

Qy 2545 accttacaataatctaaccgacaaagatcatatccatgtgtatgtgttagacatttt 2604

Db 2217 ACTTTATCAATAATCTAACCCAGCAAGATCATATCCATGTGTATGTGTAGACATTTT 2276

Qy 2605 attcattgactaaaccagacagtttcagtgtacaaattgtgtccctctgttcagc 2664

Db 2277 ATTTCAATTGACTAACCCAGCAGATTTCAGTGATGCAAATTTGTGCCCTCTGCTCAGC 2336

Qy 2665 tgaacacagtcctggaattcaaaaaccttgataaagtcctccacagttgtataaattga 2724

Db 2337 TGAACAGTCCTGACTTCAAAAACCTTGAATAAGTCCACAGTTGTATAAATTGGA 2396

Qy 2725 caattaggaattttaaccttagatgatcatattgtgtccatttttatttatttatt 2784

Db 2397 CAATTTAGGAATTTTAAACTTTAGATGATCATTTGGTTCCATTTTATTTCAATTTTAT 2456

Qy 2785 ttgtttaatgcaaacaggaacttaaatgaacttgcctctgtttttaaagattataaaaa 2844

Db 2457 TTTGTTAATGCAACAGGACTTAAATGAACCTTGATCTCTGTTTAAAGATTATTTAAAAA 2516

Qy 2845 acattgtgtatctatacatatggtccttgaggactagcttccactacacagatat 2904

Db 2517 ACATTGTGTATCTATACATATAGGCTCTTGAGGACTTACCTTTCACTACACTACAGGATAT 2576

Qy 2905 gatctcatgtagtcacataaaccctgcagagtgatattccagagtgctcgatactgtta 2964

Db 2577 GATCTCATGTAGTCCATATTAACCTGCAGAGTGATTTTCCAGAGTGCTCGATCTGTTA 2636

Qy 2965 attacatctcatagtaggctggaagaagaatgacctagcttctgtatatacagctgttgc 3024

Db 2637 ATTACATCTCCATTAGGGCTGAAGAAGATGACCTACGTTTCTGTATACAGCTGTGTGCT 2696

Qy 3025 ttctgattgtgttactgtacacagagaagtgtgtgcactgaggtctgcgtgtgtccgta 3084

Db 2697 TTTGATGTGTGTACTGTACACAGAAGTGTGTGCACCTGAGGCTCTGCGTGTGCTCCGTA 2756

Qy 3085 tggaaaaacctgtagccctgcaggttaagtaactgtctccattcatgtttacgtggaat 3144

Db 2757 TGGAAAGCCTGTAGCCCTGCGAGTTAAGTACTGCTTCATTGTTAGCTGGAAT 2816

Qy 3145 ttctctcccatggaatgttaagtaaaacttaagtgttgcatcataaataatgttaact 3204

Db 2817 TTTTCTCCCATGGAATGTAAAGTAAAACTTAAGTGTTCATCAATAAATGTAATACT 2876

Qy 3205 aaa 3207

Db 2877 AAA 2879

RESULT 7

AL133174/c AL133174 113836 bp DNA PRI 14-JUN-2000

LOCUS Human DNA sequence from clone RP3-470L14 on chromosome 20, complete

DEFINITION sequence.

ACCESSION AL133174

VERSION AL133174.15 GI:8573761

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 113836)

AUTHORS Ramsay,H.

TITLE Direct Submission

JOURNAL Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT requests: clonerequests@sanger.ac.uk

On Jun 20, 2000 this sequence version replaced gi:8388434.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr20

RP3-470L14 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-470L14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP3-470L14 is at 113836 in this sequence. The true right end of clone RP1-155G6 is at 100 in this sequence.

FEATURES

source 1..113836

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/clone="RP3-470L14"

/clone_lib="RPCI-3"

3061..3077

/note="Weak data"

9583..95627

/note="Single clone region. pUC subcloned to give 12x coverage. Assembly confirmed by restriction digest data."

BASE COUNT 33974 a 25115 c 24233 g 30514 t

ORIGIN

Query Match 42.2%; Score 1357.4; DB 34; Length 113836;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1838 gtgtggaggtgtctgaaccttctctgcatgaaccattataaaatcccaacatatatc 1897

Db 34965 GTGTGGAGGTGCTGAACCTTTCTGCGCATGAACCATTTATAAATCCCAACATATATAC 34906

Qy 1898 tgaataactgaactgtcttgaaaaatttgaattctgtataacctccagtgggccgagag 1957

Db 34905 TGAATAACTGAAACTGCTTGAATAATTGGAATTTCTGATACCTCCAGTGGCCGAGAG 34846

Qy 1958 acacggtggtaagatgtggcagcagcaggygaagacacagaaacacaaggagcg 2017

LYOVELSVGGQOENGKTRQAKHDAAKALRILONEPLPERLEVNGRESEENLNK
SEISOVEEIALKRNLPVNFVARESGPBMKNFVTKVSVGEFVGEGEKSKISKNA
AIAVLEELKLPLPAVERVPRIKKTPIVKPOTSPEYOGGINPISRLAIOQAKK
EKEPEYTLTLTREGLPRRREFYMQVKVGNHTAEGTGNKKVAKRNAENMLELGFVP
QRSHQTRTOVRGEDTHKETRMWKSNI,FLNLALGMENTSNKEDFRMPYLSHQLP
GILPMVPEVAQAVVSQGHKDFTRAAPNPAKATVTAMIARELLYGTSPTAETILK
NNISGHVPHGPLTRPSEOLDYLSRVQGFQVEYKDFPKNNKNEFVSLINCSQPLIS
HGIGKDVESCHDMAALNIKLSELDOOSTEMPRITGNPMSVCGRC"
BASE COUNT 891 a . 693 c 699 g 783 t
ORIGIN

Query Match		87.0%;	Score 2799.4;	DB 68;	Length 3066;
Best Local Similarity		99.5%;	Pred. No. 0;		
Matches 2870;		Conservative 0;	Mismatches 6;	Indels 7;	Gaps 6;
QY	328	gcataaccctactgtagaactaaatgcactgtgcatgaaacttggaaaaaaccatgt	387		
Db	1	GCATAACCCCTACTGTAGAACTAAATGCACCTGTCATGAAACTTGAAAAAACCAATGT	60		
QY	388	ataagcctgttgacccttactctcgatgcaqtcacactataactacaatagagag	447		
Db	61	ATAAGCCTGTGACCCCTTACTCTCGATGCGACGCCACCTATACTACACATGAGAGAG	120		
QY	448	gtgcttatcccccggagtacttltaccatltccagtlccaccttacttatacaagt	507		
Db	121	GTGCTTATCCCCCGAGTACTTTTACCCATTTCAGTTCACACTTTACTTTATCAAGTGG	180		
QY	508	aacttctgtgagagagacagcaatltaatgcaaaaggaagaaagacagagctgcgaac	567		
Db	181	AACTTTCTGTGGAGGAGACGCAATTTAATGGCAAGGAAGACACAGACAGGCTGCCAAAC	240		
QY	568	acgatgctgtgcgaacglttagatcctgcagaaltgagccctgccagagagctgtg	627		
Db	241	ACGATGCTGTGCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCCTGCCAGAGAGGCTGG	300		
QY	628	aggtgaatggaagaaatccgaagaagaatactcaataaactgaaataagltcaag	687		
Db	301	AGGTGAATGGAAGAAATCCGAAGAAAMCTCAATAATCTGAATAATGAAGTCAAGTGT	360		
QY	688	ttgagattgcacttaacggaacttgcctgtgaatttcgagtgcccgaggagagtgcc	747		
Db	361	TTGAGATTGCACCTTAACGGAACCTTGCCCTGTGAATTTTCGAGGTGGCCCGGAGAGTGGCC	420		
QY	748	caccccacatgaagaactttgtgaccaaagttlcgggttgggagtttctgtygggaggtg	807		
Db	421	CACCCACATGAAGAACTTTGTGACCAAGGTTTCGGTTGGGAGTGTGTGGGGGAAGGTG	480		
QY	808	aaaggaaaaagcaagaagatttcaaaaaaatgcgcacatagctgttctttaggagctga	867		
Db	481	AAGGAAAAAGCAAGAAGATTTCAAAAAATAATGCCCATAGCTGTCTTGAGGAGCTGA	540		
QY	868	agaagttaccgccccctgcctgcagttgaacgagtaaaagcctagaatcaaaaaagaacaa	927		
Db	541	AGAAGTTACCGCCCCCTGCCTGCAGTTGAACGAGTAAAGCCTAGAAATCAAAAAAGAAACAA	600		
QY	928	aaccatagtcgaagccacagacaagcccaataatgccaagggatcaatccgattagcc	987		
Db	601	AACCCATAGTCAAGCCACAGACAAGCCAGAAATATGGCCAGGGATCAATCCGATTAGCC	660		
QY	988	gactggccccagatccagcagggcaaaaaaggaagagagccagagtacacgctcctcaca	1047		
Db	661	GACTGGCCCCAGATCCAGCGCAAAAAAGAGAGAGAGCCAGAGTACACGCTCTCACAG	720		
QY	1048	agcgagcctcccgccgagcagggagttgtgatgcaagltgaaggttgaaaccacactg	1107		
Db	721	AGCGAGGCTCTCCCGCGCCGAGGAGTTGTGATGCAGGTTGAAGGTTGGAACACACACTG	780		
QY	1108	cagaaggaaacggygcaccaacaagaaggtgyccaagcgaatgcaagccgaagaacatgctg	1167		
Db	781	CAGAAAGGAACGGGACCAACAAGAAAGGTGGCCCAAGCGCAATGACGCCGAGACAATGCTGG	840		
QY	1168	agatccttggttccaagtcccgcaagcgcagcccaaccaaacccgcactcaagltcagag	1227		

Db	841	AGATCCTTGTTTCAAGTCCCGCAGCG - -CAGCCACCAACCCGCACTCAAGTCAGAGG	898		
QY	1228	agaagacaccccataaagaacccagggatggaagaagaagtaacc - tttttgaacctg	1286		
Db	899	AGAAGACACCCATTAAGAAACCAAGGGGATGGAAGAAAGTAACCTTTTGTGAACCTGGC	958		
QY	1287	tctgggat - gaaatggactagtaataaagaagatgagltcagagatgcctataag	1345		
Db	959	TCTGGGATGGAATATGGACTAGTAATAAAGAGGATGAGTTCAGGATGCTTATCTAAG	1018		
QY	1346	tcatacagcagctgcctgtgnaattcttcccatgylgcccgaagtcgcccagctgtag	1405		
Db	1019	TCATCAGCAGCTGCCTGCTGGAATTTCTCCATGCTGCCGAGGTGCCAGGCTGTAG	1078		
QY	1406	agttagtcagagacatcacaccaagaatttaccagggcagctccgaatccctgcgaagc	1465		
Db	1079	AGTTAGTCAAGACATCACACCAAGATTTTACAGGGCAGCTCCGAATCCTGCCAAGGC	1138		
QY	1466	caagtaactgcacatagaccgagagttgtgtatgtygggcacctcgcacacagccga	1525		
Db	1139	CACGGTAACCTGCCATGATAGCCCCGAGAGTGTGTATGGGGGCACCTCGCCACAGCCGA	1198		
QY	1526	gaaccattlaaagaataacatctctcagggcaacgltaccccatgaaacctcacaagacc	1585		
Db	1199	GACCATTTTAAAGAAATACATCTCTTACGGCCAGCFACCCCATGACCTCTCACAGAGAC	1258		
QY	1586	ctctgagcaactgacatcttccagagtcgaagatltccaggttgaaatacaaaagctt	1645		
Db	1259	CTCTGAGCAACTGACATATCTTCCAGAGTCAGGGATTCCAGGTTGAATACAAAGACTT	1318		
QY	1646	ccccaaaaacaagaacgaatttgtatctctatacaattgtcctcctcagccaactct	1705		
Db	1319	CCCCAAAAACAACAAGAAAGAAATTTGTATCTCTTATCAATTTGCTCCTCACACACTCT	1378		
QY	1706	gatacagcatggtatcggaagatgtgtgagtcctcgcatgatatggtctgcgtgaacat	1765		
Db	1379	GATCAGCCATGTTATCGGCAAGGATGTGAGTCTCCATGATATGCTCGCGCTGAACAT	1438		
QY	1766	cttaaaagttgctgtcgaaglttgacaacaagaagltacagagatgccaagaacaagaaacg	1825		
Db	1439	CTTAAAGTTGCTGTGAGTTGGACCAACAAGTACAGAGATGCCAAGAAGAAACGG	1498		
QY	1826	accaatgtctgtgtgtyggaggtgtgtaacccttcttgccatgaaacatataaaatcc	1885		
Db	1499	ACCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAACCATTTATAAATCC	1558		
QY	1886	caacatatatactgnaaataactgnaactgtcttgnaaaatttgnaatttctgatacctca	1945		
Db	1559	CAACATATATACTGAATAATACTGAAACTGCTTGAAATAATTGGAATTTCTGATACCTCA	1618		
QY	1946	gtgggcccagagacacaggttggtaagatgtgtyggcagcagcagggagaacaacaagaac	2005		
Db	1619	GTTGGCCGAGAGACACGGTGGTTAAAGGATGTGGGCCAGCAGCAGGGAGACACAAGAAAC	1678		
QY	2006	acaaggaagcggtgtggcc - ggtggaactgtgtggtgtgtgtgtgtgtgtgtgtgtgtgt	2064		
Db	1679	ACAAGGAGGGCGGTGTGGCCGGGCTGGACTGTGGGGGGGTTGTGTGTGTGTGTGTGTGT	1738		
QY	2065	tgaactgycgttccctacgcaatlagcagctgtcctgttgggaagaaggtgtgccagccag	2124		
Db	1739	TGACCTGGCGGTCTTACGCAATAGCAGCTGCTGTGGGAAGAGGGGCTGCCAGCCAG	1798		
QY	2125	ctggtctcccggaacacagcagatccacacccctgggcacctccgtgtgtgtgtgtgt	2184		
Db	1799	CTGTTCTCCCGGACACCAAGATCCACACCTGGGACCTCCGTGTTGTTGTTGTTGTTT	1858		
QY	2185	ttccctgtgtgaagaagaacagcgcagcacccttcaagctgtgtcactcagacac	2244		
Db	1859	TTTCCCTGTGTGAAGAAGAAACGGCAGCAGCCCTTCTCAAGCTGTGCTCACTCAGACAC	1918		
QY	2245	atgggacaaacccctgacagccatgtccagagagagccttgcagggcccagagctaa	2304		

QY	1632	gaalacaaagactlccccaaaaaacaagaacgaattgtlatcctlatcaatgtctcc	1691
Db	1921	GAATACAAAGACTTCCCAAAAAACAACAAGAACGAATTTGTATCTTATCAATTTGCTCC	1980
QY	1692	tctcagccacctctgtatcaagccatgtgtatcgcgaagatctgtgagtcctgcgatgatgtg	1751
Db	1981	TCTCAGCCACCTCTGATCATGCCCATGTGTATCGCAAGAGATGTGAGAGTCCCTGCATGATATG	2040
QY	1752	gctgcgctgaacatctttaaagttgtctgtctgtgaqtltgacccaacaagtacagagatgcc	1811
Db	2041	GCTGCGCTGAACATCTTTAAAGTTGCTGCTGTGAGTTGGACCAACAAGTACAGAGATGCCA	2100
QY	1812	agaacacggaacgcgacccaatgtctgtgtgtggaggtgcctgaaccttctcggccatgaa	1871
Db	2101	AGAACAGGAACGACCAATGCTGTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAA	2160
QY	1872	ccattataaaatccccacaatalactgaanaatactgaanaactgtcttgaanaatttgaat	1931
Db	2161	CCATTATAAATCCCAACATATATACTGAAAATACTGAACCTGCTTTGAAAATTTGCAAT	2220
QY	1932	ttctgataccctccagtgtygcgagagacacagtgtygtgtaaaagatgttgycagcagcaggg	1991
Db	2221	TTCTGATACCTCCAGTGGGCCGAGAGACACGCGTGGAAGATGTGGCCACGACGAGGG	2280
QY	1992	aagacacacagaacacacaaagagcgcgctgtgycgcgctgtygactgtgtctygggttctgt	2051
Db	2281	AAGACACACAGAAACACACAGAGCGCGCTGTGGCCGGCTGGACTGTGCTGGGGTTGTGT	2340
QY	2052	gatggccaactcgtgtgacctgycggtccctacgcataagcagctgcctgttgggaaagag	2111
Db	2341	GATGGCCACTCGGTGACTGGCGGCTGCCCTACGCAATAGCAGCTGCCCTGTGGGGAAGAGG	2400
QY	2112	gctgcccagccagctgtgtctcccgagacaccagcagatccacacccctgycgacctccgt	2171
Db	2401	GCTGCCCAGCCAGCTGTGTCTCCCGGACACCAGCAGATCCACACCCCTGGGCACCTCCGT	2460
QY	2172	gtttggtcttttttccctgtgtgaagaagaacgycacgacccctctcaagctgg	2231
Db	2461	GTTTGGCTCTTTTTCCTCCCTGTGTGAAGAAGAAACGGCACGACCCCTTCTCAAGCTGG	2520
QY	2232	ctcactcagacacatctgggacaaacccctygacagccaatgccaagagagagccttgaaccg	2291
Db	2521	CTCACTCAGACACATTTGGGACAAACCCTGGACAGCCCATGCCAGAGAGAGGCGTTTGACCG	2580
QY	2292	gccccagagctaaaagcaccagagaaaaatcaaatgtctctactcagcgtgacccaactt	2351
Db	2581	GCCCCAGAGCTAAAGCACACAGAAAAATCAAAATGCTTCTACTCAGCGTGACCCAACTT	2640
QY	2352	tctcagtgtyccacggccccaccacccctcctgcagtaaccaacacacatcaccactgtctct	2411
Db	2641	TTCTAGTGTGCCACGGCCCCACCACTCCTGCAGTACCACACCATCACCTGCCTTCT	2700
QY	2412	cttccaacagtgatctgtatctcttagtttcattattcttcttttgattgatgatgacatat	2471
Db	2701	CTTCCAACAGTGATCTGTATCTTAGTTCATTTATTTCTTTTGATGTGATATGACACTAT	2760
QY	2472	ataaaatttcaatcttgaagaatttctcaatgtlatctcagttaaaatagcacaglttggaaac	2531
Db	2761	ATAAAATTTTCATTTGAGAAATTTCTCAATGTATCTAGTTAAATAGCACAGATTGGAAAC	2820
QY	2532	ttgtctgagactgacttatacaataatcaaccgacaaagatcalatccatgtgtatgtg	2591
Db	2821	TTGTCTGAGACTGACTTTATCAATAATCTAACCAGCAAAAGATCATATCCATGTGTATGTG	2880
QY	2592	gttagacattttatctcatctgactaaccacgacaglttcagtgatgcaaatgtgtgc	2651
Db	2881	GTTAGACATTTTATTTTCATTTGACTAACCACGACAGTTTCAGTGATGCAAAATTTGTGC	2940
QY	2652	cctctggttcagctgaacaacagtcctgtgacttccaanaaccttgaataagtcctccacagt	2711
Db	2941	CCTCTGTTCAGCTGAACAACAGTCTCTGGACTTTCAAAAACCTTGAATATAGTCTCCACAGT	3000

QY	2712	tglataaatlgyacaatttaggaattttaaactltagatgatcatltygtlccatl	2771
Db	3001	TGTATAAATTGGACAAATTAGGAATTTAAACTTTAGATGATCATTTGGTTCATTTT	3060
QY	2772	ttlcaattlbatlbtgttaaltgcaaacagagacttaaaatgaactlgtactgtl	2831
Db	3061	TTTCATTTTATTTTGTGTTAATGCAAAACAGGACCTTAATGAACCTTGATCTCTGTTTAA	3120
QY	2832	agatlataaaaaacatltgtlatctatacatatggtcctltaggagcttagctt	2891
Db	3121	AGATTATTAAAAAACATGTGTATCTATACATATGCGCTTGAGGACCTTAGCTTTC	3180
QY	2892	cactacagagatatgatctccatgttaglccatataaacctgcagagltgatl	2951
Db	3181	CACCTACAGGATATGAMCTCCATGTAGTCCATATFAAACCTGCAGAGTGATTTTCCAGAGTG	3240
QY	2952	ctcgatactgtltaatacatctccatltagggctgnaaaagaatgacctagct	3011
Db	3241	CTCGATACTGTTAATATACATCTCCATTAGGGCTGAAGAATGACCTACGTTTCTGTATA	3300
QY	3012	cagctgtgttgccttltgatgtlgtlctactgtacacagaagtgtlgtcactg	3071
Db	3301	CAGCTGTGTGCTTTTGATGTGTGTACTGTACACAGAAAGTGTGTGCACCTGAGGCTCTG	3360
QY	3072	cgtgtgtccgtatgaaaaaacctgtgtagccctgcaglttaagtlactgct	3131
Db	3361	CGTGTGTCGGTATGAAAAACCTGTAGCCCTCGGAGTTAAGTACTGCTTCCATTCATTG	3420
QY	3132	tttaagctgnaattlctctcccatggaatgtgaagtaaaaaactlaagtlt	3191
Db	3421	TTTACGCTGGAATTTTCTCCCATGGAATGTAAGTAAACCTTAAGTGTGTGTCATCAAT	3480
QY	3192	aaatgylaatactaaaaa	3217
Db	3481	AAATGGTAATACTAAAAA	3506

```

RESULT: 6
LOCUS      HSA132258
DEFINITION Homo sapiens mRNA for stauflen protein, partial.
ACCESSION  AJ132258
VERSION    AJ132258.1  GI:4572587
KEYWORDS   stauflen gene; stauflen protein.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 3066)
AUTHORS   Falcon,A.M., Fortes,P., Marlon,R.M., Beloso,A. and Ortin,J.
TITLE     Interaction of Influenza virus NS1 protein and the human homologue
            of Stauflen in vivo and in vitro
JOURNAL   Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE   99263025
REFERENCE  2 (bases 1 to 3066)
AUTHORS   Ortin,J.
TITLE     Direct Submission
JOURNAL   Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,
            Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN
FEATURES
source     .
            1..3066
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             36..1526
              /gene="stauflen"
              36..1526
               /gene="stauflen"
               /function="dsRNA binding protein"
               /codon_start=1
               /product="stauflen protein"
               /protein_id="CAB40082.1"
               /db_xref="GI:4572588"
               /translation="MKLGKKPMYKPVDPYSRMOSTYNYNMRGAYPPRYFPFPVPL

```

/translation="MSQVQVQVONPSALSQSQILNKNSLISQPLMSIPSTSSLP
ENAGRIPIONSALPSASITSTSAAESITPTVELNALCMKLGKKPMKPVDPYSRMOST
YNYNMRGAYPPRYFPFPVPLPYQVELSVAGQOFNGKGTROAKHDAAMALRIL
QNEPLPERLEVNGRESEENLNKSEISQVFEIALKRNLPNFEVARESGBPKNFVT
KVSVEVEGESEKSKISKKNAAIAVLEELKLPPLPAVERVKPRIKKTKPIVKPQ
TSPEYGGQINPISRLAQIQAKKEPEYTLTERGLPRREFVMQVKVGNHTAEGTG
TNKKVARNNAENMLEILGFVPOQOPTKPAKSEKTPDKPGDGRKVTPEEGSD
ENGTSNKEDFRMPYLSHQILPMPVEVAQAVQSQGHHTKDFTRAAPNPAKAT
VTAMIARELLYGGTSPTAETILKNNISGHPHGPLFRSEQLDYSRVQGFQVEYKD
FPKNKNKNEFVSLINCSSQPLISHIGIKDVESCHDMAALNILKLSELDQOSTEMPRT
GNPMSVCGRC"

BASE COUNT 965 a 875 c 790 g 876 t
ORIGIN

Query Match 90.6%; Score 2914.8; DB 10; Length 3506;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 3215; Conservative 0; Mismatches 2; Indels 289; Gaps 1;

QY 1 actcctgcgagctgcgagcgctgagcgctcttcagcgcttgcgagcgagctgcgct 60
Db 1 ACTTCCTCGCGGCTGCGGCGCTGAGCGCTTTCAGCGTTTGCGGCGGCGCTGCGCGT 60
QY 61 ctctctcgagctcgagctctcttgacgcgctcccccccgagcgccgagcgagcgctc 120
Db 61 CTCTCTCGGCTCGCGCTCTTTCAGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCTC 120
QY 121 ctccagcgccactccgcctctccctccctcctcctcctcctcctcctccttctcctt 180
Db 121 CTCACGCGCCACTCGCGCTCTCCCTCCCTCGTCCTCTTCTCTCCCTTTTTCCTT 180
QY 181 ctctctccctcctcctcgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
Db 181 CTTCCTTCCTCTCTCGCGCGCCACCGCCAGAGCGCGCGCGCGCGCGCGCGAGC 240
QY 241 agcagccagagttatataccacttaacctctcagaactgacaagaacacattgtcc 300
Db 241 AGCAGCCAGAGTTATTATACCACTTAACCTCTCAGAACTGAACAAGACACATTGTCC 300
QY 301 tggaaagccctctttaaataa----- 322
Db 301 TGGAAAGCCCTCTTTTAAAAAGGTAGAACTTTAGACTTCATAGCACTGAATTAACCTG 360
QY 323 ----- 322
Db 361 CACTGAAAGCTGTTTACCTGCATTGTTTACATTTGTTGAAGTGACCATGTCTCAAGTT 420
QY 323 ----- 322
Db 421 CAAGTGCAAGTTCAGAACCATCTGCTCTCTCAGGGAGCCAAATACTGAACAAGAAC 480
QY 323 ----- 322
Db 481 CAGTCTCTTCTCTCACAGCCTTGATGAGTATTCCTTCTACTACTAGCTCTGCCCCCT 540
QY 323 ----- 322
Db 541 GAAAAATGCAGGTAGACCCATTCAAAACTCTGCTTTAACCTCTGCATCTATTACATCCACC 600
QY 323 -----agaaagcataacccctactgtagaactaaatgcactgtgcatagaact 371
Db 601 AGTGCAGCTGCAGAAAGCATAAACCCCTACTGTAGAACTAAATGCACCTGCATGAACCTT 660
QY 372 ggaataaaacccaatgtataagcctgttgaccccttactctcgatgcagtcacactataac 431
Db 661 GGAATAAAACCAATGTATAGCCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATAAC 720
QY 432 tacaacatgagagaggtgtctatacccccagagttacttaccacattccagttccacct 491
Db 721 TACAACATGAGAGAGGTGCTTATTCGCCCGAGGTTACTTTTACCCATTTCAGTTCCACCT 780
QY 492 ttactttatcaagtggaaacttctgtggagagacagcaatttaatgcaaaaggaagaca 551
Db 492 ttt 551

Db 781 TTACTTTATCAAGTGTGAACCTTCTGTGGGAGAGACCAATTTAATGCAAGAAAGACA 840
QY 552 agacaggtcgcgaacacacgatgtctgtgcacaaagcgttgagatcctgcagaatgagccc 611
Db 841 AGACAGGCTGCGAAACACGATGCTGCTGCCAAAGCCTTGAGGATCTGCAGATGAGCCC 900
QY 612 ctgcagagagagctgtgaggtgaaatgaaagagaatccgaagaagaanaatcctaataatc 671
Db 901 CTGCGCAGAGAGGCTGAGGTGAATGGAAGAGAAATCCGAAGAAGAAATCTCAATAATCT 960
QY 672 gaaataagtcgaagtgtttgagattgacacttaacgcgaactgtcctgtgaatttcgagtg 731
Db 961 GAAATTAAGTCAAGTGTGAGATTGCACCTTAACGGAACCTTGCCCTGTGAATTTGAGGTG 1020
QY 732 gcccgagagagtgcccaacccacatgaagaacttgtgacaaagtttcggttgaggag 791
Db 1021 GCCCGGAGAGTGGCCCAACCCACATGAAGAACCTTTGTGACCAGGTTTCGTTGGGAG 1080
QY 792 ttgttgaggagagtgaaaggaagaaagcaagaagattcaagaanaatgcgcacatagct 851
Db 1081 TTTGTGGGGGAGAGGTGAAGGGAAGAAAGCAAGAAGATTTCAAAGAAAAATGCCGCATAGCT 1140
QY 852 gtcttgagagagctgaagaagtlaccgccccctgcctgcagtgtaagcagtaga 911
Db 1141 GTTCTTGAGAGAGCTGAAGAAGTTAACCGCCCCCTGCCTGCAGTTGAACGAGTAAAGCCTAGA 1200
QY 912 atcaaaaagaacaaacccatagtlcaagccacagacaagccacagatatgycagagg 971
Db 1201 ATCAAAAAGAAAACAAACCCATAGTCAAGCCACAGACAAAGCCAGATATGGCCAGGGG 1260
QY 972 atcaatccgattagccgacttggccagatccagcagycacaaagaagaaagagagccagag 1031
Db 1261 ATCAATCCGATTAGCCGACTGGCCACAGATCCAGCAGGCAAAAAAGAGAGAGCCAGAG 1320
QY 1032 tacaagctcctcacagagcgaagcctcccgcgccgagggagtttgtatgcaggtgaag 1091
Db 1321 TACACGCTCTCACAGAGGAGGCGCTCCCGCGCGCAGGAGTTGTGATGCAGGTGAAG 1380
QY 1092 gtggaaaaccacactgcagaagaaagcggcaccacaagaaggttggccaagcgcaatgca 1151
Db 1381 GTTGGAAACACACACTGCAGAAAGAAAGGCGCACCAACAAGAGGTGGCCAAAGCAATGCA 1440
QY 1152 gccgagaacatgtctgagatccttgtttcaaaagtcgccgagcgagcccaaccaaaccc 1211
Db 1441 GCCGAGAACATGCTGAGATCCTTGTTTCAAGTCCCGCAGCGGACGCCAACCAACCC 1500
QY 1212 gcactcaagtcagagagaagacacccataaagaacacagggatgtgaagaaagtaacc 1271
Db 1501 GCACCTCAAGTCAGAGGAGAGACACCCATTAAGAAACCAAGGGATGGAAGAAAGTAACC 1560
QY 1272 tttttgaacctgtctgtggatgaaatgagactagtaataagagatgagttcagag 1331
Db 1561 TTTTGTGAACCTGCGCTGTGGGATGAAATGGGACTAGTAATAAGAGAGATGAGTTCAGG 1620
QY 1332 atgccttatctaagtcacagcgtgcctgtcgtgaattcttcccatgtgcccaggtc 1391
Db 1621 ATGCCCTTATCTAAGTCATCAGCAGCTGCCCTGCTGGAATTTCTCCCATGTGCCCGAGGTC 1680
QY 1392 gcccaagctgtagaggttaagtcgaagacatcacaccaaagattttaccaaggcgagctcg 1451
Db 1681 GCCCAGGCTGTAGGAGTTAGTCAAGGACATCACACCAAGATTTTACCAGGGCAGCTCCG 1740
QY 1452 aatcctgcacaagggcacaagtaactgcatgtagcccgagagttgtgtatggyggcacc 1511
Db 1741 AATCCTGCCAAGGCCACGGTAACTGCCATGATAGCCCGAGAGTTGTGTATGGGGCACC 1800
QY 1512 tcgcccacagccgagacattttaaagaataacatctctcagggccaagtlaccacatgca 1571
Db 1801 TCGCCACAGCCGAGACCATTTTAAAGAAATAACATCTCTTCAGGGCCACGTACCCCATGGA 1860
QY 1572 cctctcaagagaccctctgagcaactgagctatcttccagaggtccagggatccaggt 1631
Db 1861 CCTCTCACGAGACCCCTGTGAGCAACTGACTATCTTCCAGAGAGTCCAGGATTCACAGGTT 1920

Db 1546 GATTCAGGTTGAATACAAAGACTTCCCCAAAAACACAGACGAATTTGTATCTCTTA 1605
QY 1681 tcaattgctcctctcagccaaccttgatcagccatggtatcggcaagatgtgagtcct 1740
Db 1606 TCATTTGCTCTCTCAGCCACCCTCTGATCAGCCATGGTATCGCAAGATGTGAGTCTT 1665
QY 1741 gccatgatatgctgcgtggaacatctaaagttgctgtgagttggaaccaaaagta 1800
Db 1666 GCCATGATATGGCTCGCGTGAACATCTTAAAGTTGCTGCTGAGTTGACCACAAGTA 1725
QY 1801 cagagatgccaaagaacagaaacccaatgtctgtgtgaggagtgctgaaccttt 1860
Db 1726 CAGAGATGCCAAGAACAGAAACGACCAATGCTGTGTGGGAGGTGCTGAACCTTTT 1785
QY 1861 ctggccaatgaaacctatataaatcccaacatatatactgaaanaactgacttga 1920
Db 1786 CTGGCCATGAAACCATTTATAAATCCCAACATATATACTGAAATACTGAACCTGCTTGA 1845
QY 1921 aaatttggaatttctgataacctcgaagtggcgagagacacggtggttaaagatgtgg 1980
Db 1846 AAATTTGGAATTTCTGATACCTCCAGTGGGCCGAGAGACGCGTGGGTAAAGGATGTGG 1905
QY 1981 cagcagcagggaaagacaacagaaacacaaggagcgctgtgcccgtctgactgtcgt 2040
Db 1906 CAGCAGCAGGGAAAGACAACAGAAACAAAGAGCGCGCTGTGGCCGCTGACTGTGCTG 1965
QY 2041 gggttgtgtgatgagccactcgtgtgacctgagcggtccctacgcaatagcagctgctgt 2100
Db 1966 GGGTTGTGTGATGGCCACTCGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCTGT 2025
QY 2101 ggggaagaagggtctgccacgacgctgtctctccgggacacccagcagatccacacctg 2160
Db 2026 GGGGAAGAGGGGCTGCCACGCCAGCTGGTCTCCCGGGACACCAGAGATCCACACCTTG 2085
QY 2161 ggcacctccgtgttgtgtcttttttccccctgtgtgaaagaagaaacggcaccgacct 2220
Db 2086 GGCACCTCGCTGTTGGTCTTTTTCCTCTGTGTGAAGAAGAAACGGCACGACCCCT 2145
QY 2221 tctcaagctggctcaactcagacacatgtggacaaacctgagacagccatgccagaagag 2280
Db 2146 TCTCAAGCTGGCTCACTCAGACACATTTGGGACAAAACCTGGACAGCCATGCCAGAGAG 2205
QY 2281 gcccttgaccggccccagagctaaagcaccagagaaatcaaatgcttctactcagcg 2340
Db 2206 GCCTTTGACCGGCCCCAGAGCTAAAGCACACAGAAAAATCAAAATGCTTCTACTACGCG 2265
QY 2341 tgaccaaaccttctagtgtgccaaggccccaccacctcctgcagtlacccacacatcac 2400
Db 2266 TGACCCCACTTTCTAGTGTGCCACGGCCCCACACCTCCTGCAGTAGCCACACATCAC 2325
QY 2401 cactgcttctcttcgaacagtgatctgtatctctagtltcattatcttcttcttgatga 2460
Db 2326 CACTGCTTCTCTTCCAACAGTGAATCTGTATTCTTAGTTCAATTATTTCTTTTGATTGA 2385
QY 2461 tatgaactataaaaaatttcaattgagaattctcgaatgtatctagttaaatagcac 2520
Db 2386 TATGACACTATATAAAATTTTCAATTGAGAAATTTCTCAATGTATCTAGTTAAATAGCAC 2445
QY 2521 agttggaacctgtctgagactgacttatacaataatcaaccgacaagaatcatatcc 2580
Db 2446 AGTTTGGAACTTGTCTGAGACTGACTTTATCAATTAATCTAACCGACAAGATCATATCC 2505
QY 2581 atgtgtatgtgttagacattttatttcatltgactaacccagagacagtttcagtgtgc 2640
Db 2506 ATGTGTATGTGGTTAGACATTTTATTTCATTGACTAACCCAGACAGAGTTTCAGTGTGC 2565
QY 2641 aaattgtgtgccctctgtgttcagctgaaacagtcctcgtgactttcaaaaaccttgaataag 2700
Db 2566 AAATTTGTGCCCTCTGTGTTACGCTGAACAGAGTCCCTGGACTTTCAAAAACCTTGAAATAG 2625
QY 2701 tctcccaagttgtataaaattggaacaatttaggaattttaaacctttagatgatcatctgg 2760
Db 2626 TCTCCACAGTGTGTATAAATTTGACAATTTAGGAATTTTAAACTTTAGATGATCAATTTGG 2685

QY 2761 ttccattttatttcaatttttatttcttgaatgcaaacaggaactlaatgaacttgat 2820
Db 2686 TTCCATTTTATTTTCATTTTATTTTGTAAATGCAAAACAGAGACTTAAATGAACTTGAT 2745
QY 2821 ctctgttttaagaattataaaaaacatgtgtatctatataatgctcttgagacct 2880
Db 2746 CTCTGTTTAAAGATTATTAAAAAACATTTGTGTATCTATACATATAGGCTCTTGAGAGCTT 2805
QY 2881 agctttcaactacactacagagatatgtctccatgtagtgtccatatataaacctgcagagtgc 2940
Db 2806 AGCTTTCACCTACACTACAGGATATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGAT 2865
QY 2941 ttccagagtgtcgcatactgttaattacatctccatlaggctgaaaagaatgacctac 3000
Db 2866 TTTCAGAGTGCCTGCATACCTGTTAATTACATCTCCATTAGGCGCTGAAAAAATGACCTAC 2925
QY 3001 gttctgtatacagctgtgtgtcttctgattgtgtgttactgtacacagaagtgtgca 3060
Db 2926 GTTCTGTATACAGCTGTGTGCTTTGTGATGTTGTGTACTGTACACAGAGAAGTGTGCA 2985
QY 3121 tccattcatgtttaacgctgtgaatttttctcccatatggaatgtaagtaaaacttaagt 3180
Db 3046 TCCATTCAATCTTTACGCTGGAATTTTCTCCCATATGGAATGTAAGTAAACTTAAGTGT 3105
QY 3181 ttgtcatcaataaatgtaataactataaaaaaa 3217
Db 3106 TTGTCAATCAATTAATGTTAATACTTAAAAA 3142

RESULT 5
AF061939
LOCUS AF061939 3506 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061939
VERSION AF061939.1 GI:4335948
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3506)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3506)
AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source location/Qualifiers
1..3506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1..3506
/gene="STAU"
409..2142
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AAD17531.1"
/db_xref="GI:4335949"

TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada

FEATURES
source
1. .3142
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. .3142
/gene="STAU"
288. .1778
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="stau protein"
/protein_id="AADI7532.1"
/db_xref="GI:4335951"
/translation="MKLGKKPMYKPPDYSRMSTYNNMRGAVPPRYFYPPEVPVPL
LYQVELSVGGQGFNGKGTROAKHDAAKALRILONEPLPERLEVNGRESEENLNK
SEISQVEFIALKRNLPVNFVARESPGPHMKNFVTKVSQVGEFVGEQSKISKNA
AIAVLEELKLPPLPAVERVKPRIKKTPIVKPOTSPEYGGQINPLSLAQIQAKK
EKEPEYTLITERGLPRRRFVQVYKVGHTAEGTGNNKVAKRNAENMLEILGFVP
QROPTKPAKLSSEKTPIKKPGDGKVTFFEPGSGDNGTSNKEDEFMBYLSHQLPA
GILPMVPEVAQAVGSQGHHTKDFTRAAPNPAKATVTAMIARELLYGGTSPTAETILK
NNISGHPHGLPTRPSEQDLYLSRVGFQVEYKDEPKNNKNEFVSLINCSSQPLLIS
HGIGKDVESCHDMAALNLIKLSLDDQOSTEMPTGTNGPMSVCGRC"

BASE COUNT 862 a 778 c 733 g 769 t
ORIGIN

Query Match	94.98;	Score 3053.8;	DB 10;	Length 3142;
Best Local Similarity	97.68;	Pred. No. 0;		
Matches 3140;	Conservative	0;	Mismatches 2;	Indels 75; Gaps 1;
QY 1	acttcctgcgaggtcgcgagcgccctgagcgctctcagcggttgcgcgcgctgcgct	60		
Db 1	ACTTCCTGCCGGGCTGCGGGCGCTGAGCGCTCTTACGCGTTTGCGGGCGGCTGCGCGT	60		
QY 61	ctctctcgctcccgcttcccttgaccgcctcccccccgcccgcgcgcccgctc	120		
Db 61	CTCTCTCGGCTCCCGCTTCCTTTGACCGGCTCCCCCCCCCGGGCGGCGGCGGCTC	120		
QY 121	ctccacggccactcgcgctcttccctccctcgctccttcttctcctccttcttctt	180		
Db 121	CTCCACGGCCACTCGGCTCTTCCTCCCTTCGTCCTTCCTTCCTTCCTTCCTTCCT	180		
QY 181	cttccttccctcctcgcgcgccaacgcgcccagagccgcgcccggggagagctcgagc	240		
Db 181	CTTCCTTCCTTCCTTCGCGCCACCGCCAGAGCCGCGCGGGGGAGAGTCCGGAGC	240		
QY 241	agcagccagagttattaaccaacttaacctcagaaactgaaacaagaacattgttcc	300		
Db 241	AGCAGCC-----	247		
QY 301	tggaacgcccctcttttaaaaaaagaagcataaccctactctgtagaactaaatgcactgt	360		
Db 248	-----AGAAAGCATTAACCCCTACTTTAGAACTAAATGCACCTGT	285		
QY 361	gcatgaaacttgaaaaaaacccaatgtataagcctgttgaccctactctcgcgatgcagt	420		
Db 286	GCATGAAACTTGGAATAAACCAATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGT	345		
QY 421	ccaacctataactacaacatgagagggaggtgcttatcccccgaggtacttltaccatttc	480		
Db 346	CCACCTATACTACACATGAGAGAGGTGCTTATCCCCCGAGGTACTTTTACCATTTC	405		
QY 481	cagttccacccttactcttatcaagtlygaacttctgtgagagacagcaaatataatgca	540		
Db 406	CAGTTCACCTTTACTTTATCAAGTGAACCTTCTGTGGGAGGACAGCAATTTAATGGCA	465		
QY 541	aaagaaagacaagacagctgcgaaacacgaatgcctgcctgcctgcgcaaaqcqtltagatcctgc	600		

Db	466	AAGCAAGACACAGACAGCGCTGCCGAACACAGCATGCTGCTGCCAAAGCGTTGAGGATCCTGC	525
QY	601	agaatgagccctgcccagagaagcttggagtgtaatggaagaagaatccgaagaagaatc	660
Db	526	AGAAATGAGCCCCCTGCCAGAGAGGCTGGAGGTGAATGGAAGAGAAATCCGAAGAGAAATC	585
QY	661	tcaataaatctgaataaagtcgaagtgtttagatgtaactttaaacygaacttgcctgtga	720
Db	586	TCAATAAATCTGAATAATAGTCAAGTGTGTGAGATTGCACTTAAACGGAACCTTGCTGTGA	645
QY	721	atttcgaggtgccccggagagagtgccccaccaccacatgaagaacttgtgaccgaagttt	780
Db	646	ATTTCGAGGTGGCCCGGAGAGTGGCCACCACCACATGAAGAACCTTGTGACCAAGGTTT	705
QY	781	cggltggggagtttgtggyggaagtgaaaggaaagcaagaagattccaagaaaatg	840
Db	706	CGGTTGGGGAGTTGTGGGGGAAGGTGAAGGGAAGCAAGAAGATTTCAAAGAAAAATG	765
QY	841	ccgcatagctgttctcttgagagagctgaagaagttaccgccccctgcctgcagttgaaag	900
Db	766	CCGCCATAGCTGTTCTTGAGAGAGCTGAAGAAGTTACCGCCCCCTGCCTGCAGTTGACGAG	825
QY	901	taaagcctagaatcacaagaaaaacaaaccccatagttcaagccacagacaagccagaat	960
Db	826	TAAAGCCTAGAAATCAAAAAAGAAAAACAACCAACCATAGTCAAGCCACAGACCAAGCCAGAAAT	885
QY	961	atggccaggggatacaatccgaattagccgaactgccccagatccagcagcaaaaaaggaga	1020
Db	886	ATGGCCAGGGGATCAATCCGATTAGCCGACTGGCCCAATCCAGCAGGCAAAAAAGGAGA	945
QY	1021	aggagccagagtacacgctcctcacagagcgagcgctcccgcgccgagggagttgtga	1080
Db	946	AGGAGCCAGAGTACACGCTCTCTCACAGAGCGAGCGCTCCCGCGCGCAGGAGGATTGTGA	1005
QY	1081	tgcaggtgaaggttggaaaccacactgcagaaggaacgggcaaccaagaaggtggcca	1140
Db	1006	TGCAGGTGAAGGTTGGAAACCACTGCAGAGAAGGAACGGGCAACCAAGAAGGTGGCCA	1065
QY	1141	agcgcaatgcagccgagaacatgctgtgagatccttggtttcaagtcccgcgagcgagc	1200
Db	1066	AGCGCAATGCAGCCCGAGAGACATGCTGGAGATCCTTGGTTTCAAAAGTCCCGCAGCGGACG	1125
QY	1201	ccaaccaaacccgcactcaagtcagagagagaagacacccataaagaaaaccagggatggaa	1260
Db	1126	CCACCAAAACCCGCACCTCAAGTCAAGAGGAGAGAGAACACCATTAAAGAAACGAGGGATGAA	1185
QY	1261	gaaaagtaaccttttttgaacctgctctggtgatgaanaatggagtaataaagag	1320
Db	1186	GAAAAGTAACCTTTTGAACCTTGCTCTGGGATGAAAATGGGACTAGTAATAAGAGG	1245
QY	1321	atgagtcagagatgccttatctaagtcatacagcagctgcctgtgnaattcttccatgg	1380
Db	1246	ATGAGTTCAGGATGCCCTTATCTAAGTCATCAGCAGCTGCCCTGGAAATCTTCCCATGG	1305
QY	1381	tgcgcgaggtgcgccagagctgtgagagttagtcraagacatcacaccaagaatttaca	1440
Db	1306	TGCCCGAGGTCGCCCCAGGCTGTAGAGATTAGTCAAGGACATCACACCAAGATTTTACCA	1365
QY	1441	gggcagctccgaatcctctgccaagccaagtgtaactgccaatagtagccgagagttgtgt	1500
Db	1366	GGGCGAGCTCCGAATCCTGCCCAAGGCCACGGTAACTGCCATGATAGCCCCGAGAGTTGTGT	1425
QY	1501	atggggcaactcgccacacagccgagaacattttaaagaataacatctcttcagggcacg	1560
Db	1426	ATGGGGCACCTCGCCACACAGCCGAGACCATTTTAAAGAAATAACATCTCTTCAGGGCCAG	1485
QY	1561	taccocatgacctctcacagagaccctctgagcaactgtaatacttccagagttccag	1620
Db	1486	TACCCCATGGACCTCTCACGAGACCTCTGAGCAACTGGACTATCTTCCAGAGTCAGG	1545
QY	1621	gattccaggttgatacaagacttccccaaaaacaacaagaagaaatttgtatatctta	1680

QY 1369 tcttcccaatggtgcccgaagtcgcccagcgtgtagagtltagtcaagacatcacacca 1428
|||||
Db 1501 TTCTTCCCATGGTGCCCGAGGTCGCCAGGCTGTAGAGTTAGTCAAGGACATCACACCA 1560
QY 1429 aagatttaccaggcagctccgaatccctgcgaagcccaagcgttaactgcatgtatagccc 1488
|||||
Db 1561 AAGATTATACCAGGCGAGCTCCGAATCCTGCCAAGGCCACGGTAAC TGCCATGATAGCCC 1620
QY 1489 gagagtgtgttatgtggygcacccctgcgccacagcccgagaccattttaagaataacatct 1548
|||||
Db 1621 GAGAGTTGTGTATGGGGGCACCTCGCCACAGCCGAGACCATTTTAAGAAATTAACATCT 1680
QY 1549 cttcaggccaagtaaccacatggaacctctcacgagaccctctgagcaactgatacttt 1608
|||||
Db 1681 CTTCAAGCCACGTACCCCATGGAACCTCTCACGAGACCCTCTGAGCAACTGACTATCTTT 1740
QY 1609 ccagagtcagggaatccaggttgaaatacaaaactccccaaaaacaacaagaacgaat 1668
|||||
Db 1741 CCAGAGTCCAGGGATTCAGGTTGAATACAAAGACTTCCCAAAAAACAACAAGAACGAAT 1800
QY 1669 ttgtatctctatcaattgtctctctcagccacctctgatcagccatgtgtatcgccaag 1728
|||||
Db 1801 TTGTATCTCTTATCAATGTGCTCTCTCAGCCACCTCTGATTCAGCCATGATGGCAAG 1860
QY 1729 atgtgagtgctcgtccatgatatgtgctgctgaacatcttaagtgtgtctgagttgg 1788
|||||
Db 1861 ATGTGAGTCCCTGCCATGATATGCGCTGCGCTGAACATCTTAAAGTTGCTGTCTGAGTTGG 1920
QY 1789 accaacaagaagtaacagagatgccaaagaacaggaacgaacaaatgtctgtgtggyaggt 1848
|||||
Db 1921 ACCAACAAAGTACAGAGATGCCAAGAACAGAACGAAACCAATGTCTGTGTGGGAGGT 1980
QY 1849 gctgaaccttctctgycatgaacacaltataaaatcccaacatatatactgaaataactg 1908
|||||
Db 1981 GCTGAACCTTTTCTGGCCATGAACCATTTATAAATCCCAACATATATACTGAAATAACTG 2040
QY 1909 aaactgcttgaanaatttgnaatttctgataacctcagctggyccgagagacaggtggt 1968
|||||
Db 2041 AAACGCTTTGAAAATTGGAATTTCTGATACCTCCAGTGGGCCGAGAGACAGCGTGGGT 2100
QY 1969 aaagatgtgycagcagcagcagggaaagacaacaagaanaacaagagcgctgtgcygc 2028
|||||
Db 2101 AAAGATGTGGGACAGCAGCAGGAAGACAACAGAACAGGAGGCGGCTGTGGCCGGC 2160
QY 2029 tggactgtctggyggttctgtgtgatgycgaactcgtgtgacctgycggtccctaagcaata 2088
|||||
Db 2161 TGGACTGTGCTGGGGTTTGTGTGATGGCCACTCGGTGACCTGGCGGTCCCTTACGCAATA 2220
QY 2089 gcagctgcctgtggygaagaagcgctgcccagcagcgtgtctcccggyacacagcag 2148
|||||
Db 2221 GCAGCTGCCTGTGGGAAGAGGGCTGCCAGCAGCGTGGTTCTCCCGGACACCGAGCAG 2280
QY 2149 atccaacacctggygcacctcggtgttgctcttttcccccgtgtgtgaagaagaac 2208
|||||
Db 2281 ATCCACACCCCTGGGCACCTCGTGTGGTCTTTTTCCTTCCCTGTGTGAAGAAGAAGAAC 2340
QY 2209 ggcacgaccccttctcaagctggtctactcagacacacattggygacaacccctgagacca 2268
|||||
Db 2341 GGCACGACCCCTTCTCAAGCTGGCTCACTCAGACACATTGGGACAAACCCCTGGACAGCCA 2400
QY 2269 tgcagaagaagcgcttgcaccgccccagagctaaagaacacagaganaaatcaaatgct 2328
|||||
Db 2401 TGCCAGAGAGAGGCGCTTGACCGGCCCCAGAGCTTAAAGCACACAGAGAAAATCAAAATGCT 2460
QY 2329 tctactcagcgtgacccaacttttctagtgtgcccaggyccccaaccactcctgcagtac 2388
|||||
Db 2461 TCTTACTCAGCGTGACCAACTTTTCTAGTGTGCCAGGGCCCCACACACCTCCTGCAGTAC 2520
QY 2389 ccacaccatcaccaactgtctctcttccaacagtgatctgtatctctagtctcatattt 2448
|||||
Db 2521 CCACACCATCACCACTGCTTCTCTTCCAAACAGTGATCTGTATTCTTAGTTTCAATTATT 2580

QY 2449 tcttttgattgatatgacacctataataaatttccatttgagaatttctcaattgtatcta 2508
|||||
Db 2581 TCTTTTGATTGATATGACACTATATAAAATTTTCATTGTGAGAATTTCTCAATTTGATCTA 2640
QY 2509 gttaaatagcacaglttggaaacctgtctgagactgacttatacaataatcaaccgaca 2568
|||||
Db 2641 GTTAAATAGCACAGTTTGGAAACTTGTCTGAGACTGACTTTATCAATATATTAACCGACA 2700
QY 2569 aagatcatatccatgtgtatgtgttagacatttlatcaltgactaaccagacag 2628
|||||
Db 2701 AAGATCATATCCATGTGTATGTGGTTAGACATTTTATTTCAATTGACTAACCCAGGACAG 2760
QY 2629 ttccagtgatgcaaatltgtgcccctcgtgttcagctgaaacagtcctgacttcaaaa 2688
|||||
Db 2761 TTTCAGTGTATGCAAAATGTGTGCCCTCTGTTCAAGCTGAACAGTCCGTGACTTCAAAA 2820
QY 2689 acctgataaagtcctccacagttgtataaatgtgacaatttaggaattttaacttag 2748
|||||
Db 2821 ACCTGATAAAGTCTCCACAGTTGTATAAATTTGACAAATTTAGGAATTTTAACTTTAG 2880
QY 2749 atgataattgtgtccatttlatcatttlatcatttltgttaagtgaacaggaacttaa 2808
|||||
Db 2881 ATGATCATTTGGTTCATTTTATTTTCATTTTATTTTGTATATGCAACAGACACTTAA 2940
QY 2809 atgaacttgaatcctgttlttaagaaltlaaaaaacattgltatcatalatagtc 2868
|||||
Db 2941 ATGAACTTGATCTCTGTTTAAAGATTATTAAAAACATGTGTATCTATACATATAGGC 3000
QY 2869 tcttgagacttagcttccactaactacagagatatgatctccatgtagtccataaanaac 2928
|||||
Db 3001 TCTTGAGGACTTAGCTTTCACATACACTACAGGATATGATCTCCATGTAGTCCATATAAAC 3060
QY 2929 ctgcagagtgatlttccagagtgctcgataactgtttaatcatatccattagggctgaaa 2988
|||||
Db 3061 CTGCAGAGTGATTTTCCAGAGTGCCTCGATACTGTATAATTACATCTCCATTAGCGTGAAA 3120
QY 2989 agaatacctaagcttctgtatacagcgtgtgtcctttagatgtgtgtactgtacaca 3048
|||||
Db 3121 AGAATGACCTACGTTTCTGTATACAGCTGTGTGCTTTGTGATGTTGTACTGTACACA 3180
QY 3049 gaagtgtgcaactgagcgtcgtcgtgtgtgtccgtatgysaaaacctgtgtagccctgcag 3108
|||||
Db 3181 GAAGTGTGTCACTGAGGCTGTGCGTGTGTCCGTATGGAACCACTGTAGCCCTGCGAG 3240
QY 3109 ttaagtaactgttccattcatattgtttaagcgtggaatttctcccatggaatgaagta 3168
|||||
Db 3241 TTTAAGTACTGCTTCCATTCATGTGTTAACGCTGGAATTTTCTCCCATGGAATGTAAGTA 3300
QY 3169 aaacttaagtgttgtcatcaataaatgtaataactaaaaa 3217
|||||
Db 3301 AAACTTAAGTGTGTGTATCAATATAATGGTAATACTTAAAAA 3349

RESULT 4
AF061940
LOCUS 3142 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061940
VERSION AF061940.1 GI:4335950
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3142)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE Mammalian stauflen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3142)
AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3349)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3349)
AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1..3349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1..3349
/gene="STAU"
495..1985
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AAD17533.1"
/db_xref="GI:4335953"
/translation="MKLGKKPMYKPPVDPYSRMQSTYYNMRGAYPPRYFPVPPL
LYQELSVGGQFNGKGRTRQAAKHDAAKALRLIONELPERLEVNGRESEENLNK
SEISQVEIALKRLNLPVNFVARESGPRPHMKNFTKVSVEFVGESEKSKISKNA
AIAVLELKKLPLPAVERVKPRIKKTKPIYKQTSPEYGGINPISRLAQIQAKK
EKEPEYTLTERGLPRREFVMQKVGNHTAEGTGNKKAARNAENMLEILGKVP
QROPKPAKSEKTPIKKPGDKRVTFEEDSGDENGTSNKEDEFKMPYLSHQQLPA
GILPMPEVAQAVGVSQGHHTKDFTRAAPNPAKATVTAMIARELILYGSTPAETILK
NNISSGHVPHGPLTRPSEOLDYLSRVQGFQVEYKDFPKNNKNEFVSLINCSSQPLIS
HGIGKDVESCHDMAALNLKLSELDOQSTEMPRGTNGPMSCVGRG"
BASE COUNT 913 a 838 c 777 g 821 t
ORIGIN
Query Match 95.5%; Score 3071.8; DB 10; Length 3349;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 3215; Conservative 0; Mismatches 2; Indels 132; Gaps 1;
QY 1 actctcgccggctgaggcgccctgagcgtcttcagcgtttgagcgcgcgctgagcgt 60
|||||
Db 1 ACTTCCTGCCGGCTGCGGGCGCTGAGCGCTCTTACGCTTTCGGCGGCGCTGCCGT 60
QY 61 ctctctcgctcccgctctcttgcacgcctcccccccccccgcccgcgcgcgccgctc 120
|||||
Db 61 CTCTCTCGGCTCCGCTTCTTGACCGCTCCCCCCCCCGCGCGCGCGCGCCGCTC 120
QY 121 ctccagcgccactccgctcttccctcccttcgctccttctctctctctcttcttctc 180
|||||
Db 121 CTCACGGCCACTCCGCTCTTCCCTCCCTTCCTTCCTCTCTCTCTCTCTCTTTTCTT 180
QY 181 ctctctccctctctcgccgcaacgcccagagcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
|||||
Db 181 CTTCCTTCCCTCTCTGCGCGCACCGCCAGACCGCGCGCGCGCGGAGCAGAGTCCGAGC 240
QY 241 agcagcca----- 248
|||||
Db 241 AGCAGCCAGCAGCAGCAGGTGAGTTTGTCTTGTGCGCCAGGCTGAGTGACAGTGGC 300
QY 249 ----- 248
Db 301 GTGATCTCGGCTCACTGCAACCTCCACCTCCACAGGTCAAGCATTTTCCACTTCAGCCTC 360
QY 249 ----- 288
-----gagttattaaccaacttaacctctcagaactgaacaaga 288
|||||

Db 361 CCGATAAGCTGAGATTACAGAGTATTATTAACCACTTAACCTCTCAGAACTGAACAAGA 420
QY 289 caacattgttcctggaacgcctctctttaaaaaagaagcataacccctactgtagaac 348
|||||
Db 421 CAACATGTGTCGGAACGCCCTCTTTTAAAAAAGAAAGCATTAACCCCTACTGTAGAAC 480
QY 349 taatgcaactgtgcatgaaacttggaaaaaaaccatgtatlaagcctgttgaccctact 408
|||||
Db 481 TAAATGCACCTGTGCATGAACCTTGAAAAAAACCAATGTATAGCCTGTGACCCCTACT 540
QY 409 ctcgatgcaagtcacactataactacaacatgagagagtgcttatccccgaggtact 468
|||||
Db 541 CTCGATGCAGTCCACCTATACTACACATGAGAGAGGTGCTTATCCCGAGGTAAGT 600
QY 469 ttaccatttccagttccaccttacttatacaagtgaacttctgtggagagacgc 528
|||||
Db 601 TTTACCATTTCCAGTTCACACCTTTACTTTATCAAGTGAACCTTCTGTGGAGAGACAGC 660
QY 529 aattaatgcaaggaagaacaagacagcgtgcaaacacgctgtgccaagcgt 588
|||||
Db 661 AATTTAATGCAAGGAAGAAAGACACAGACAGCTGCGAAACACGATGCTGCTGCCAAAGCGT 720
QY 589 tgagatctctgcgaatgagccctgcgaagagagcgtgaggtgtaatggaagaatccg 648
|||||
Db 721 TGAGGATCCTGCAGAAATGAGCCCCCTGCCAGAGAGCGCTGAGAGTGAATGGAAGATCCG 780
QY 649 aagaagaaaatctcaataaatctgaataatgtaagtttgagattgcacttaacgga 708
|||||
Db 781 AAGAAGAAATCTCAATAATCTGAATAATAGTCAAGTGTTCAGATTGCACCTTAACGGA 840
QY 709 acttgctgtgaattcgaagttgcccggagagagttgcccacccacatgaagaacttg 768
|||||
Db 841 ACTTGCTGTGAATTTGAGGTGGCGCCGGAGAGTGGCCACCCACATGAAGAACCTTG 900
QY 769 tgaccaagtttcggttgggagttgtgagggaaggtgaagggaaggaagcaagaagatt 828
Db 901 TGACCAAGGTTTCGGTGGGAGTTTGTGGGGGAAGGTGAAGGGAAGCAAGAAGATTT 960
QY 829 caaagaaaatgcgcgcacatagctgttcttgagagagctgaagaagttaccgccccctgctg 888
Db 961 CAAAGAAAATGCCGCCCATAGCTGTCTTGAGGAGAGCTGAAGAGATTACCGCCCTGCTG 1020
QY 889 cagttgaacgagtaaaagcctagaatcaaaaaaagaacaaaccccatagtcgaagccacaga 948
Db 1021 CAGTTGAACGAGTAAAGCCTAGAAATCAAAAAAAGAAACCAACCATAGTCAAGCCACAGA 1080
QY 949 caagcccaagaatatgcccagggagatcaatccgattagccgactgcccagatccagcagg 1008
Db 1081 CAAGCCCAAGATATGGCCAGGGGATCAATCCGATTAGCCGACTGGCCCAAGATCCAGCAGG 1140
QY 1009 caaaaaagagaagagcagagtlacacgctctccacagagcagagcctccgcgcgcga 1068
Db 1141 CAAAAAAGGAGAGAGAGCCAGAGTACACGCTCTCACAGAGCGAGGCTCCCGCGCGCA 1200
QY 1069 gggagttgtgatgcaagtlgaagttggaacacacactgcagaaggaagcgcacaca 1128
Db 1201 GGGAGTTTGTGATGACAGGTGAAGTTGGAACCAACACTGCAGAAAGCAACGGCACCAACA 1260
QY 1129 agaagtggtgccaagcgcaatgcagccgagacaatgctgagatcctgtgttcaagtc 1188
Db 1261 AGAAGTGGCCCAAGGCCAATGACGCCGAGAACATGCTGAGAGATCTTGGTTCAAAAGTCC 1320
QY 1189 cgcagcgcgcaccccaaacccgcactcaagtgcagaggaagaacacccataaagaac 1248
Db 1321 CGCAGCGGACGCCCAACCAACCCGCACTCAAGTCAAGAGGAGAGACACCCATAAAGAAC 1380
QY 1249 caggggatggaagaaaagtaacctttttgaaacctgctctgaggatgaaatgagacta 1308
Db 1381 CAGGGATGGAAGAAAGTAACCTTTTGAACCTTGCGCTCTGGGGATGAAAATGGGACTA 1440
QY 1309 gtaataaagagatgagttcagatgccttatactaatcatcagcagctgcctgtgtaa 1368
Db 1441 GTAATAAAGAGATGAGTTCAAGATGCCCTTATCTAAGTCATCAGCAGCTGCCTGTGGAA 1500

QY	1231	agacaccataaagaaccaggygatlgaagaaagtaccttlttgaacctgctctg	1290
Db	1199	AGACACCCATAAAGAACCA - GGGATGGAAGAAAGTAACCTTTTGAACCTGGCTCTG	1257
QY	1291	gggatlgaanaatggygactgaataaagagatgagttcagatgaccttatctaagtcac	1350
Db	1258	GGGATGAAAAATGGGACTAGTAATAAAGAGGATGAGTTCCAGATGCCCTATCTAAGTCATC	1317
QY	1351	agcagctgcctgctgnaattctcccatggtgcgccaggtgcgccagctgtagagtta	1410
Db	1318	AGCAGCTGCCCTGCTGGAAATTTCTCCCATGTTGCCCGAGCTCGCCACAGGCTGTAGAGTTA	1377
QY	1411	gtcaaggacatcacaccaaaagatttaccagggcagctccgaaltcctgcgaagccacg	1470
Db	1378	GTCAGGACATCACACCAAGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGG	1437
QY	1471	taactgcatgatagcccagagagtgtgtatggyggcacctcgccacagccgagacca	1530
Db	1438	TAACTGCCATGATAGCCCGAGAGTTGTTGATGGGGCACCTCGCCACAGCCGAGACCA	1497
QY	1531	tttaaagaataacatcctcttcagggccacgtlaccctcatgacctctcacgagaccttg	1590
Db	1498	TTTTAAAGATAACACCTCTTCAGGCCACGTACCCCATGGACCTCTCACGAGACCTCTG	1557
QY	1591	agcaactggaactatcttccagagatccagggatccaggtltgaatacaaaagcttccca	1650
Db	1558	AGCAACTGGACTATCTTCCAGAGTCCAGGGATTCAGGTTGAATACAAGACCTCCCCA	1617
QY	1651	aaaacaacaagaacgaattgtlalcctlatcaattgtcctctcagccacctgtalca	1710
Db	1618	AAAAACAACAAGAAAGAAATTGTATCTCTTATCAATTCCTCTCAGCCACCTGTGATCA	1677
QY	1711	gccatggtatcggcaagatgttgagtgctcctgcctatgatatgctgcgctgaacatctta	1770
Db	1678	GCCATGGTATCGGCAAGGATGTGGAGTCCCTGCCATGATATGGCTGGCGCTGAACATCTTAA	1737
QY	1771	agttgctgtctgagttggaaccaacaagtacagagatgccaagaacaggaacgaccaa	1830
Db	1738	AGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAGGAACGGACCAA	1797
QY	1831	tgctcgtgtgtggygaggtgctgaaccttctctgcccatagaaccatlataaaatcccaaca	1890
Db	1798	TGTCGTGTGTGGAGGTGCTGAACCTTTCTGCGCATGAACCATTTATAAAATCCCAACA	1857
QY	1891	tatatactgaanaatactgnaaactgcttltgaaatltgnaatttctgataacctccagtggy	1950
Db	1858	TATATACTGAAAAATACTGAAACTGCTTTGAAAAATTTGGAATTTCTGATACCTCCAGTGGG	1917
QY	1951	ccgagagacacaggttgygtaaagatgttgggcagcagcaggygaagacaacagaaacacaag	2010
Db	1918	CCGAGAGACACCGGTGGTAAGGATGTGGGCAGCAGCGGAAGACAAACAGAAAACACAAG	1977
QY	2011	gaggcgyctgtgccc - ggcctggaactgtgtcgtggygttgtgtgtagtgccactcgtgacc	2069
Db	1978	GAGGCGGCTGTGGCCGGCTGGACTGTGTGGGTTTGTGTGATGGCCACACTCGGTGACC	2037
QY	2070	tggcggtccctacgcaataagcagctgcctgttgyggaagaaggyctgccagccagctggt	2129
Db	2038	TGGCGGTCCCTACGCAATAGCAGCTGCTGTGGGGAAGAAGGCTGCCACGCCAGCTGGT	2097
QY	2130	tctcccggaacaccagcagatccacacccctgggcacctccgtgttgttcttlttcc	2189
Db	2098	TCTCCCGGACACACGAGATCCACACCCTGGGCACTCCGTGTTTGGTCTTTTTTTTC	2157
QY	2190	cctgtgtgaagaagaanaacgycacgaccccttctcaagcttgctcaactcagacacattgg	2249
Db	2158	CCTGTGTGAAGAAGAAACGGCACGACCCCTTTCAAAGCTGGCTCACTCAGACACATTGG	2217
QY	2250	gacaaacccttgacagccatgtccagagagagggcctlttgaccggccccagagctaaagca	2309
Db	2218	GACAAACCCCTGGACAGCCATGCCAGAGAGAGGCCCTTTGACCGGGCCCCAGAGCTAAAGA	2277
QY	2310	ccagagaaatcaaatgtcttctactcagcgtgaaccaacttctagttgtccacgccc	2369

Db	2278	CCAGAGAAATCAATGCTTCCTACTCAGCGTGACCCCACTTTCTAGTGTGCCACG GCC	2337
QY	2370	ccaccaccctcctgcagtlacccacacccatcacccactgcttctcttccaacagtgatctgt	2429
Db	2338	CCACCACTCCTCGAGTACCACACACCATCACCACTGCTTTCTCTCCAAACAGTATCTGT	2397
QY	2430	attcttagtltcatltaatttcttlttgattgatatgacactataaaatlbtcatlttag	2489
Db	2398	ATTCTTAGTTCAATTATTTTCTTTTGATTGATATGACACTATATAAAATTTTCATTTGAG	2457
QY	2490	aatttcacatgttatctagtltaaatagcacagtltgynaactgtctgagactgtt	2549
Db	2458	AATTTCTCAATTGTATCTAGTTAAATAGCACAGTTTGGAAACTTGTCTGAGACATGACTTT	2517
QY	2550	atcaataatctaacccgacaaaagatcatatccatgtgtatgtgttagacatlttlattc	2609
Db	2518	ATCAATATATCTAACCGACAAAGATCATATCCATGTGTATGTGTAGACATTTTATTTTC	2577
QY	2610	attgactaacccagagacttccagtgatgcaaaatgtgtgcctctgttccagctgaaa	2669
Db	2578	ATTGACTTAACCCAGAGACAGTTTCAGTGATGCAAAATGTGTGCCCTCTGGTTACGCTGAAA	2637
QY	2670	cagtcctgagacttcaaaaacctlgaataagcttcccacagltgtataaatlgacaatt	2729
Db	2638	CAGTCCCTGGACTTTCAAAAACCTTGAATAAGTCTCCCAAGTGTATATAAATTGGACAATT	2697
QY	2730	taggaattttaaactttagatgatcatlttggttccatlttlatlttcatlttcatltt	2789
Db	2698	TAGGAATTTTAAACTTTAGATGATCATTTGGTTCATTTTATTTTCATTTTACTTTTGT	2757
QY	2790	taatgcaaacagagacttaaatgaactlttgatctctgtltttaaagattatlaaaaaacatt	2849
Db	2758	TAATGCAAAACAGAGACTTAAATGAACCTTGATCTCTGTTTTAAAGATTATTAAAAACATT	2817
QY	2850	gtgatctatacatatgctctcttgaggacttagcttccactacactacagatatgatct	2909
Db	2818	GTTATCTTATACATATGGCTCTTGAGGACTTAGCTTCACTACACTACAGGATATGATCT	2877
QY	2910	ccatgtagtcataataaacctgcagagtgatlttccaagagtgctcgatactgttaattac	2969
Db	2878	CCATGTAGTCCATATTAACCTGCAGAGTGATTTTCCAGAGTGCCTCGATACTGTATATTAC	2937
QY	2970	atctccattagggctgaaaaaagaatgacctacgcttctcgtatatcacagctgtgtgctttga	3029
Db	2938	ATCTCCATTAGGGCTGAAAAAGATGACCTACGTTTCTGTATACAGCTGTGTGCTTTTGA	2997
QY	3030	tggtgtgttactgtacacagaagltgtgcactgaggtctgcgtgtgtccgatatgaa	3089
Db	2998	TGTTGTGTACTGTACACAGAAGTGTGTGCACGTGAGGCTCTGCGTGTGTCGTAATGAAA	3057
QY	3090	aacctgttagccctgcgagtttaagtactgcttccatcatgtttaaagctgnaatttct	3149
Db	3058	AGCCTGGTAGCCCTGCGAGTTAAGTACTGCTTCCATTCATTTGTTTACGCTGGAATTTTTC	3117
QY	3150	tccocatggaatgtlaagtaaaaacttlaagtgttltgtcatcaataaatggttaatactaaaaa	3209
Db	3118	TCCCCATGGAATGTAAGTAAAGACTTAAAGTGTGTGTCTCATCAATAATGTAATACTAAAAA	3177
QY	3210	aaaaaaaa 3217	
Db	3178	AAAAAAAA 3185	

RESULT	3		
AF061941			
LOCUS	AF061941	3349 bp	mRNA
DEFINITION	Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced, complete cds.		PRI 04-MAR-1999
ACCESSION	AF061941		
VERSION	AF061941.1	GI:4335952	
KEYWORDS	.		
SOURCE	human.		

QY	3121	tcacattcatgtttacgctggaattttctcccccacgtgaatgtlaagtaaaactlaagt	3180
Db	3121	TTCCATTCATTGTTTACCTGGAAATTTTCTCCCCCATGGAATGTAAGTAAACTTAAGTGT	3180
QY	3181	ttgtcatcataaataatggttaataactaaaaa	3217
Db	3181	TTGTTCATCAATAAATGTGTAATACTAAAAA	3217
RESULT	2		
LOCUS	AK025519	3194 bp	29-SEP-2000
DEFINITION	Homo sapiens cDNA: FLJ21866 fis, clone HEP02379, highly similar to AF061938 Homo sapiens staufen protein (STAU) mRNA.		
ACCESSION	AK025519		
VERSION	AK025519.1	GI:10438059	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP clone:HEP02379.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (sites) Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
REFERENCE	2 (bases 1 to 3194)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan		
COMMENT	(E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
FEATURES	NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
source	Location/Qualifiers 1.3194 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="HepG2" /cell_type="hepatoma" /clone="HEP02379" /clone_lib="HEP" /note="cloning vector PME18SFL3" 1.3194 /note="highly similar to AF061938 Homo sapiens staufen protein (STAU) mRNA" misc_feature 1.3194		
BASE COUNT	894 a	787 c	732 g
ORIGIN	781 t		
Query Match	96.5%;	Score 3105.2;	DB 34; Length 3194;
Best Local Similarity	99.4%;	Pred. No. 0;	
Matches 3169; Conservative	0;	Mismatches 13;	Indels 6; Gaps 5;
QY	33	cttcagcgtttgcgc--gcgcgctgcgcgctctctcctcaacgacacatccgccttccttgaccgcc	90
Db	1	CTTCAGCGTTTGCGCGCGCGGCTGCCGCGTCTCTCTCGGCTCCCGCTTCCTTGACCGCC	60
QY	91	ttcccccccgccgcgcgcgcgcgcgcctctcctcaacgacacatccgccttccttcctccct	150
Db	61	T-CCCCCCCCCGCGCGCGCGCGCGCGCTCTCTCCACAGGCGCACTCCGCCCTCTTCCCTCCCT	119

[illegible]

|||||
Db 901 TAAAGCCTAGAAATCAAAAAAGAAAAAACCCATAGTCAAGCCACAGACAGGCCAGAAAT 960
QY 961 atgagcaggagatcaatccgaattagccgactgcccagatccagcagcgaataaagaga 1020
Db 961 ATGGCCAGGGGATCAATCCGATTAGCCGACTGGCCCCAGATCCAGCAGGCAAAAAAGAGA 1020
QY 1021 aggaagccagatcacacgctccctcacagagcgaggcctcccgccgagggagttgtga 1080
Db 1021 AGGAGCCAGAGTACACGCTCCTCACAGACGGAGCCCTCCCGCCGCGCAGGGAGTTGTGA 1080
QY 1081 tgcaggtgaagttggaaccacacactgacagaaggaacgagccaccaacaagaagtggcca 1140
Db 1081 TGCAGGTGAAGTTGGAAAAACACACTGCAGAGAAGGAGCGGACCAACAAGAAGTGGCCA 1140
QY 1141 agcgcaatgcagccgagaacatgctgagatccttggttccaagtcccgcagcgagc 1200
Db 1141 AGCGCAATGCAGCCGAGAACATGCTGGAGATCCTTGGTTCAAGTCCCGCAGCGGACG 1200
QY 1201 ccaccaaacccgactcaagtacagagagaagacaccataaagaacccagggatgaa 1260
Db 1201 CCACCAAAACCCGCACTCAAGTCAAGAGAGAAGACACCCATAAAGAAACAGGGATGAA 1260
QY 1261 gaaaagtaacctttttgaacctgctcctggygatgaaatggygactagtaataagag 1320
Db 1261 GAAAAGTAACCTTTTGAACCTGGCTCTGGGGATGAAAATGGGACTAGTAATAAGAGG 1320
QY 1321 atgagttcagatgacctatcttaagtcacagcagctgcctgtggaattctcccatg 1380
Db 1321 ATGAGTTCAAGATGCTTATCTAAGTCAATCAGCAGCTGCTGTGGAAATTCTTCCCATGG 1380
QY 1381 tgcgcgaagtcgcccaggcgtgtagagttagtcgaagacatcacaccaagaatttacc 1440
Db 1381 TGCCCGAGGTGCCCCAGGCTGTAGAGATTAGTCAAGGACATCACACCAAGATTTTACCA 1440
QY 1441 gggcagctccgaatcctgccaagggccaagtaactgcatgatagcgagagttgtgtc 1500
Db 1441 GGGCAGCTCCGAATCCTGCGCAAGGCCGCTAACTGCCATGATGAGCCGAGAGTTGTGT 1500
QY 1501 atgggggacctcgcccacagccgagaccattttaaagaataacatctctcagccag 1560
Db 1501 ATGGGGGACCTCGCCCCACAGCCGAGACCATTTTAAAGAATAACATCTCTCAGGCCACG 1560
QY 1561 taaccatgacacctctcacgagaccccttgagcaactgactatcttccagagttccag 1620
Db 1561 TACCCCATGACCTCTCACGAGACCCCTTGAGCAACTGGACTATCTTTCAGAGTCCAGG 1620
QY 1621 gattccaggttgaatatacaagaacttccccaaaaacaacaagaacgaattgtatctcta 1680
Db 1621 GATTCCAGGTTGAATACAAAGACTTCCCCAAAAACAACAAGAACGAATTTGTATCTCTTA 1680
QY 1681 tcaattgtctctctcagccacaccttgatcagcagcatggtatcggaagagtgtgagtcct 1740
Db 1681 TCAATTGTCTCTCAGCCACCTCTGATCAGCCATGGTATCGGCAAGGATGTGAGTCTCT 1740
QY 1741 gccatgatatgctgctgctgaacatcttaaaagttgctgtctgagttggaccaacaagta 1800
Db 1741 GCCATGATATGGCTGCGCTGAACATCTTAAAGTGTGCTGTGAGTTGGACCAACAAGTA 1800
QY 1801 cagagatgcacaagaacaggaacgcgaccaatgtctgtgtgtggaagtgctgaaccttt 1860
Db 1801 CAGAGATGCCACAGACAGGAACGGACCAATGTCTGTGTGGGAGGTGCTGAACCTTTT 1860
QY 1861 ctggccatgaaccattataaaatcccaacatatatactgaaaaactgaaactgcttga 1920
Db 1861 CTGGCCATGAACCATTTATAAAATCCCAACATATATACTGAAAAATACTGAAACTGCTTGA 1920
QY 1921 aaattggaatttctgatacctccagtgggccgagagacacggtggglaaagatgtgg 1980
Db 1921 AAATTGGAATTTCTGATACCTCCAGTGGGCCGAGAGACACGCGTGGTAAAGATGTGGG 1980
QY 1981 cagcagcagggaaagacaacagaacacaaagagggcgctgtggccggctgactgtgtg 2040
|||||

Db 1981 CAGCAGCAGGGAAGACAACAGAAACACAGAGAGCGGCTGTGGCCGCGCTGACTGTGCTG 2040
QY 2041 gggtttgtgtgataggccaactcgtgtgacctgycggtccctaagcaatagcagctgcgt 2100
Db 2041 GGGTTTGTGTGATGGCCACTCGGTGACCTGGCGGCTCCCTACGCATATAGCAGCTGCCCTGT 2100
QY 2101 ggggaagaagggctgcccaagccagctggttctccggygacaccagcagatccacaacctg 2160
Db 2101 GGGGAAGAAGGGCTGCCACAGCTGCTGCCCGGGACACACAGATCCACACCCCTG 2160
QY 2161 ggcacctcgtgtgtgtcttttttccctgtgtgaagaagaacgycagaccct 2220
Db 2161 GGCACCTCCGTGTGTCTTTTTCCTGTGTGAAGAAGAAGAACGGCACGACCCCT 2220
QY 2221 tctcaagctgctcaactcagacacattggaacaaacccctgacagcccatgccaagagag 2280
Db 2221 TCTCAAGCTGGCTCACTCAGACACATTTGGGACAAACCCCTGGACAGCCATGCCAGAGAGAG 2280
QY 2281 gcctttgacggccccagaagcctaagaacccaagagaataatcaaatgtcttccactacgcg 2340
Db 2281 GCCTTTGACCGGCCCCAGAGCTAAAGACACAGAGAAAAATCAAAATGCTTCTACTCAGCG 2340
QY 2341 tgaaccaactttctagtytgcaagcccccaacctcctgcagtaaccaacatcac 2400
Db 2341 TGACCCAACCTTTCTAGTGTGCCACGGCCCCACACCACTCCTGCAGTACCCACACCATCAC 2400
QY 2401 cactgcttctctccaacagtgaatcgtatcttagtttcaattatcttcttgatga 2460
Db 2401 CACTGCTTCTCTCCAAACAGTGATCTGTATCTTAGTTTCATATTTCTTTGATTGA 2460
QY 2461 tatgacactataataatttcaatttgagaatttctcaattgtatctagttaaalagcac 2520
Db 2461 TATGACACTATATAAATTTTCATTTGAGAAATTTCTCAATGTGATCTAATAATAGCAC 2520
QY 2521 agtttgaacctgtctgagactgacttlatcaataatctaaaccgacaagaatcatatcc 2580
Db 2521 AGTTTGAAACTGTGCTGAGACTGACTTATCAATAATCTAACCCGACAAAGATCATATCC 2580
QY 2581 atgtgatgtgttagacaatttlatcttcaatgactaaaccagagacagtttcaagtatgc 2640
Db 2581 ATGTGTATGTGTTAGACATTTTATTATTTCAATGACTAACCCAGGACACTTTCAGTGATGC 2640
QY 2641 aaattgtgcccccttggttcagctgaaacagctcctgaccttccaanaacctgaaataag 2700
Db 2641 AAATTGTGCCCTGTGGTTCAGCTGAAACAGTCCTGGACTTTCAAAAACCTTGAATAAG 2700
QY 2701 tctccacagttgtataaaattgacaatttaagaaatttaaaacttaagatcatattgg 2760
Db 2701 TCTCCACAGTGTGATAAATTGGACAATTTAGGAATTTTAAACCTTAGATGATCATTTGG 2760
QY 2761 ttccattttlatcttcaatttlattttgttcaatgcaaacagagacttaaatgaactttgat 2820
Db 2761 TTCCATTTTATTTATTTTATTTTGTATTAATGCCAAACAGGACTTAATGAACCTTGAT 2820
QY 2821 ctctgtttaaagatatataaaaacatltgtatctatacatatgctcttgagacct 2880
Db 2821 CTCTGTTTAAAGATTATTAAAAAACATGTGTATCTATACATATAGCTCTTGAGACTT 2880
QY 2881 agcttcaactacacagatatgatctccatgtaglccatatataaactgcagagtgat 2940
Db 2881 AGCTTTCACCTACACTACAGGATATGATCTGCATGTAGTCCATATATAAACCCTGCAGAGTGAT 2940
QY 2941 ttccagagtgctcgatactgttaattacatctccattagggctgaaaagaatgacctac 3000
Db 2941 TTTCAGAGAGTGCCTGACTGTTAATATACATCTCCATTAGGGCTGAAAAAGATGACCTTAC 3000
QY 3001 gttctgtatacagctgtgtgtcttltgatagtgtgttactgtlacacagaagtgtgtga 3060
Db 3001 GTTCTGTATACAGCTGTGTGCTTTTGATGTGTGTACTGTACACAGAAGTGTGTGCA 3060
QY 3061 ctgaggtctgcgtgtgtgtccgtatggaanaacctgtagccctgcgagtttaagtgct 3120
Db 3061 CTGAGGCTCTGTGCGTGTGTCGATATGGAACCTGTAGCCCTGCGAGTTAAGTACTGCT 3120

22	150.8	4.7	160979	52	AC055879	AC055879 Homo sapi
23	149.6	4.7	3173	34	AK001576	AK001576 Homo sapi
24	149	4.6	332811	72	AL390198	AL390198 Homo sapi
25	133.4	4.1	194407	54	AC068845	AC068845 Homo sapi
26	125.2	3.9	2008	34	AK023314	AK023314 Homo sapi
27	119	3.7	2847	69	HSRNTTFT	X83973 H.sapiens m
28	110.2	3.4	5237	53	DROSTAUFEN	M69111 D.melanogas
29	110.2	3.4	61267	28	AC004336	AC004336 Drosophil
30	110.2	3.4	120625	38	AC017563	AC017563 Drosophil
31	110.2	3.4	267271	28	AE003800	AE003800 Drosophil
32	105.8	3.3	8409	29	AF225924	AF225924 Drosophil
33	103	3.2	8002	29	AF225925	AF225925 Musca dom
34	85.4	2.7	171591	50	AC027018	AC027018 Homo sapi
35	73	2.3	7218	81	I66494	I66494 Sequence 14
36	68	2.1	7218	81	I66494	I66494 Sequence 14
37	65.4	2.0	60009	48	AC023436	AC023436 Homo sapi
38	63	2.0	180968	62	AC078996	AC078996 Mus muscu
39	63	2.0	258550	35	AC010758	AC010758 Homo sapi
40	62.6	1.9	42099	29	CELE55A4	U67949 Caenorhabdi
41	62.4	1.9	141892	47	AC023197	AC023197 Mus muscu
42	62.2	1.9	89669	69	HSJ340H11	AL109755 Human DNA
43	61.4	1.9	77127	36	AC011801	AC011801 Homo sapi
44	61.2	1.9	92709	37	AC015716	AC015716 Homo sapi
45	61	1.9	80725	34	AL137078	AL137078 Human DNA

ALIGNMENTS

RESULT	1					
LOCUS	AF061938	3217 bp	mRNA	PRI	04-MAR-1999	
DEFINITION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.					
ACCESSION	AF061938					
VERSION	AF061938.1	GI:4335946				
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Homo sapiens					
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	1 (bases 1 to 3217)					
TITLE	Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and Desgroseillers,L. Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum					
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)					
MEDLINE	99147057					
REFERENCE	2 (bases 1 to 3217)					
AUTHORS	Desgroseillers,L., Wickham,L. and Luo,M.					
TITLE	Direct Submission					
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada					
FEATURES	Location/Qualifiers					
source	1. 3217					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/chromosome="20"					
	/map="20q13.1"					
gene	1. 3217					
	/gene="STAU"					
CDS	363. 1853					
	/gene="STAU"					
	/function="RNA binding protein"					
	/note="alternatively spliced transcript"					
	/codon_start=1					
	/product="staufen protein"					
	/protein_id="AAD17530.1"					
	/db_xref="GI:4335947"					
	/translation="MKLGKRPWKPVDPYSRMQSTNYNMRGAYPPRYFPYPPVPL					
	LYVELSVGGQOFNGKGTROAAKHDAAKALRILONEPLPERLEVNGRESEENLNK					
	SEISOVEEIALKRNLPVNFVARESGBPMMKNFVTKVSGEEVGESEKISKNA					
	AIAVLEELKKLPLPAVERVYKRIKKTKPIVKPQTSPEYGGGINPISRLAQIQAKK					
	EKPEYTLTLTERGLPRRREFVMQVKNHTAEGTGNKKVAKRNAENMELIGFKVP					

ORQPTKPAKSEKTPIKKPDGRKVTFFEPGSGDENGTSNKEDEFMRPYLSHQOLPA
GILPMVEVAQAVGSQGHHTKDETRAAPNPAPKATYAMIARELYGTSPTAETILK
NNISGHPHGPLTRPSEQLDILSRVQGFQVEYKDFPKNNKNEFVSLNCSQPLIS
HGIGKDVESCHDMAALNLIKLSLSDQOSTEMPRGNGPMVCGRG"

BASE COUNT 889 a 796 c 742 g 790 t

Query Match 99.9%; Score 3213.8; DB 10; Length 3217;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3215; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	acttcctgcgagctgcgagcgctgcgctcttcagcgcttgcgagcgctgcgct	60
Db	1	acttcctgcgagctgcgagcgctgcgctcttcagcgcttgcgagcgctgcgct	60
QY	61	ctctctcgctccgctcttccttgaccgctctcccccccgccgagcgcgccgctc	120
Db	61	ctctctcgctccgctcttccttgaccgctctcccccccgccgagcgcgccgctc	120
QY	121	ctccacgagcactcgcgctcttccttccttcgcttccttccttccttccttcct	180
Db	121	ctccacgagcactcgcgctcttccttccttcgcttccttccttccttccttcct	180
QY	181	cttccttccttccttcgcccacgagcgccgagcgccgagcgagcgagcgagc	240
Db	181	cttccttccttccttcgcccacgagcgccgagcgccgagcgagcgagcgagc	240
QY	241	agcagccagagttatataccacttaacctcagaacttgaaacaaagacaacattgtcc	300
Db	241	agcagccagagttatataccacttaacctcagaacttgaaacaaagacaacattgtcc	300
QY	301	tggaacgccccctctttaaataaagaagacataacccctactgtagaactaatgca	360
Db	301	tggaacgccccctctttaaataaagaagacataacccctactgtagaactaatgca	360
QY	361	gcatgaaacttgaaaaaaaccaaagtataagcctgttgaccctactctcgatgcagt	420
Db	361	gcatgaaacttgaaaaaaaccaaagtataagcctgttgaccctactctcgatgcagt	420
QY	421	ccacctatactacaacatgagagagtgcttatcccccgagtgactttaccattc	480
Db	421	ccacctatactacaacatgagagagtgcttatcccccgagtgactttaccattc	480
QY	481	cagttccaccttacttatcaagtgyaactttctgtgygagacagcaattaatgca	540
Db	481	cagttccaccttacttatcaagtgyaactttctgtgygagacagcaattaatgca	540
QY	541	aaggaagacaagacagcgtcggaaacacgagtctgctgccaagcgttgagatcctgc	600
Db	541	aaggaagacaagacagcgtcggaaacacgagtctgctgccaagcgttgagatcctgc	600
QY	601	agaatgagccccctgcgagagagcgtgaggtgaaatggaagagaatcgaagaanaatc	660
Db	601	agaatgagccccctgcgagagagcgtgaggtgaaatggaagagaatcgaagaanaatc	660
QY	661	tcaataaatctgaaataagtcaagtgtttgagatgtgcacttaacggaacttgctgtga	720
Db	661	tcaataaatctgaaataagtcaagtgtttgagatgtgcacttaacggaacttgctgtga	720
QY	721	attcgaggtgccccgagagagtgccccacccacatgaagaactttgtgaccaa	780
Db	721	attcgaggtgccccgagagagtgccccacccacatgaagaactttgtgaccaa	780
QY	781	cggttgagagttgtggyggaagtgaaggygaaagcaagaagaattcaagaanaatg	840
Db	781	cggttgagagttgtggyggaagtgaaggygaaagcaagaagaattcaagaanaatg	840
QY	841	ccgcataagctgttcttgagagcgtgaagaagttaccgccccctgcctgcagttgaacgag	900
Db	841	ccgcataagctgttcttgagagcgtgaagaagttaccgccccctgcctgcagttgaacgag	900
QY	901	taaagcctagaatcaaaaaaagaaacaaacccatagtcgaagccacagacaagccagaat	960

AUTHORS Kiebler,M.A., Hemraj,I., Verkade,P., Kohrmann,M., Fortes,P.,
Marion,R.M., Ortin,J. and Dotli,C.G.
TITLE The mammalian staufen protein localizes to the somatodendritic
domain of cultured hippocampal neurons: implications for its
involvement in mRNA transport

JOURNAL J. Neurosci. 19 (1), 288-297 (1999)
MEDLINE 99088098

FEATURES Location/Qualifiers
source 1. .360
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_lib="Rat brain"
/dev_stage="embryonic"
1. .119
/gene="staufen"
<1. .>119
/gene="staufen"
/codon_start=2
/evidence=experimental
/product="Staufen protein"
/protein_id="CAA09037.1"
/db_xref="GI:4138428"
/translation="YFYPFPVPLLYQVELSVGGQDFNGKGKMRPPVKHDATA"

gene
CDS

BASE COUNT 109 a 70 c 105 g 76 t
ORIGIN

Query Match 9.1%; Score 287; DB 74; Length 360;
Best Local Similarity 87.5%; Pred. No. 4e-56;
Matches 314; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 390 tactttaccattccagttccaccttacttatacaagtgaacttctgtggagga 449
|||||
Db 2 TACTTTTACCATTCCAGTTCCACCGTTACTCTACCAAGTTGAGCTCTCCGTGGAGGA 61
OY 450 cagcaattaatgcaagaagaacaagacagctgcgaacacagatgctgcgcaaa 509
|||||
Db 62 CAGCAGTTTAATGGGAAGGAAGATGAGACCACTGTGAACATGATGCCACTGCCCGA 121
OY 510 gcgttgagatcctgcagaatgagccctgcagagagagctgtgaggtgaatgaaagaa 569
|||
Db 122 GCCCTGAGGACTCTGCAGAGTGAGCCCCCTACCAGAGAGGCTAGAGGTAATGGAAGAA 181
OY 570 tccgaagaagaataatcaataatctgaataagtcagtgtttgagattgcactaaa 629
|||
Db 182 TCAGAAGAAGAAACCTTAATAATCAGAAATTAAGCCAAGTGTGAGATTGCGCTGAAG 241
OY 630 cggaaactgcctgtgaatttcgagtgcccgagagagtgccccacaccacatgaagaac 689
|||||
Db 242 CGGAATTTGCCCTGTGAATTTTGAGGTGGCCGGGAGAGTGCCCGCACACATGAAGAAC 301
OY 690 ttgtgaccaaggttcggttggtggagttgtggggggaagtgaaaggaaagaagaa 748
|||||
Db 302 TTTGTGACCACAGGTTTCAGTTGGGAGTTTGTAGGGGAAGGAGGAAGGAAAGCAAGAA 360

Search completed: April 5, 2001, 08:30:54
Job time: 77499 sec

RESULT 13
AL158850/c
LOCUS AL158850 192597 bp DNA HTG 06-SEP-2000
DEFINITION Homo sapiens chromosome 6 clone RP3-522B19, *** SEQUENCING IN
PROGRESS ***, 11 unordered pieces.
ACCESSION AL158850
VERSION AL158850.4 GI:9800143
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 192597)
REFERENCE Direct Submission
AUTHORS Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
TITLE CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
JOURNAL requests: clonerequest@sanger.ac.uk
On Aug 13, 2000 this sequence version replaced gi:9212635.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj522B19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 12% of reads
Sequencing vector: plasmid; L08752; 87% of reads
Chemistry: Dye-terminator Big Dye; 77% of reads
Chemistry: Dye-primer-amerham; 22% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 185625 bases at least Q40
Consensus quality: 188533 bases at least Q30
Consensus quality: 190005 bases at least Q20
Insert size: 191597; sum-of-contigs
Insert size: 144116; 9.6% error; agarose-fp
Quality coverage: 5.71x in Q20 bases; sum-of-contigs Quality
coverage: 8.21x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 13837: contig of 13837 bp in length
* 13838 13937: gap of 100 bp
* 13938 17476: contig of 3539 bp in length
* 17477 17576: gap of 100 bp
* 17577 20005: contig of 2429 bp in length
* 20006 20105: gap of 100 bp
* 20106 32070: contig of 11965 bp in length
* 32071 32170: gap of 100 bp
* 32171 39772: contig of 7602 bp in length
* 39773 39872: gap of 100 bp
* 39873 43471: contig of 3599 bp in length
* 43472 43571: gap of 100 bp
* 43572 51795: contig of 8224 bp in length
* 51796 51895: gap of 100 bp
* 51896 178838: contig of 126943 bp in length
* 178839 178938: gap of 100 bp
* 178939 181919: contig of 2981 bp in length
* 181920 182019: gap of 100 bp
* 182020 184231: contig of 2212 bp in length
* 184232 184331: gap of 100 bp
* 184332 192597: contig of 8266 bp in length.
Location/Qualifiers

```
source 1. 192597
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="6"
        /clone="RP3-522B19"
        /clone_1lb="RPCI-3"
        1. 13837
        /note="assembly_fragment:02383
        fragment_chain:1"
        misc_feature 13938. 17476
        /note="assembly_fragment:01080
        fragment_chain:1"
        misc_feature 17577. 20005
        /note="assembly_fragment:02434
        fragment_chain:1"
        misc_feature 20106. 32070
        /note="assembly_fragment:03060
        fragment_chain:2"
        misc_feature 32171. 39772
        /note="assembly_fragment:00945
        fragment_chain:2"
        misc_feature 39873. 43471
        /note="assembly_fragment:00013
        fragment_chain:2"
        misc_feature 43572. 51795
        /note="assembly_fragment:02225
        fragment_chain:3"
        misc_feature 51896. 178838
        /note="assembly_fragment:01767
        fragment_chain:3"
        misc_feature 178939. 181919
        /note="assembly_fragment:01637"
        misc_feature 182020. 184231
        /note="assembly_fragment:02408"
        misc_feature 184332. 192597
        /note="assembly_fragment:02960"
BASE COUNT 58324 a 39515 c 38313 g 55431 t 1014 others
ORIGIN
Query Match 12.2%; Score 383.2; DB 70; Length 192597;
Best Local Similarity 73.2%; Pred. No. 3.3e-78;
Matches 586; Conservative 0; Mismatches 203; Indels 11; Gaps 7;
QY 2102 gtcttttttccctgtgtgaaagaagacgacgacccctctcaagctgtcac 2161
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60919 GTAATTTTATTGGCTGTGCAAAAAGAAATGAGACAGAAAGCATTCCTGCTCAC 60860
QY 2162 tcagacacattgagacaaccctgacagccatgcagagagagagccttgaccgccc 2221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60859 ACTGCTGACTCAAC-ACCTTGCTATTAACATGCTGGAGAAAGCCTTAGCTGCCCTC 60801
QY 2222 agagctaaagcaccagaaaatcaaatgcttcctactcagcgtgacccaactttcta 2281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60800 AAAAGCTGAAAAACACCGGGGAAATTAACCTGCTTCTACTCGTTACAACATTAACCTTCTA 60741
QY 2282 gtgtgccaagggcccaaccctcctgcagtgaccacaacatcaccactgttctctcc 2341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60740 GCATGCTACGG-CCAGTCACTTACTGCAGTACCACACCATTAACCACTGCTTCTCTCT 60682
QY 2342 aacagtgatcgtatctctagtttcatatttcttcttgatgatatgacactataaa 2401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60681 AATAGTGATACATATTTTAACTTTATTA-TTCTTTTGATTGAATGATGCCATATGAA 60623
QY 2402 atttcatttgaatctctcaattgtatctagttaaatagcacagtttggaaactgtc 2461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60622 AATTCATTGAAAGTTTCTCAGTTGTATCTAGTTACATAGTACAGTTTCAGAAACCTGTC 60563
QY 2462 tgagactgacttatacataatcaaccgacaagaatcatatccatgtgtatgttag 2521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60562 CCAGACTGACTTATATCAGGAATCTAACTGCAATGATCATATCCATGATATACGTGTTGT 60503
QY 2522 acattttattcatctgactaaccagacagtttcagtgatgcaaatgtgtgcctct 2581
```

```

* 53039 59129: contig of 6091 bp in length
* 59130 59229: gap of 100 bp
* 59230 65896: contig of 6667 bp in length
* 65897 65996: gap of 100 bp
* 65997 89427: contig of 23431 bp in length
* 89428 89527: gap of 100 bp
* 89528 95785: contig of 6258 bp in length
* 95786 95885: gap of 100 bp
* 95886 102762: contig of 6877 bp in length
* 102763 102862: gap of 100 bp
* 102863 110899: contig of 8037 bp in length
* 110900 110999: gap of 100 bp
* 111000 121002: contig of 10003 bp in length
* 121003 121102: gap of 100 bp
* 121103 133048: contig of 11946 bp in length
* 133049 133148: gap of 100 bp
* 133149 145755: contig of 12607 bp in length
* 145756 145855: gap of 100 bp
* 145856 165006: contig of 19151 bp in length
* 165007 165106: gap of 100 bp
* 165107 188243: contig of 23137 bp in length
* 188244 188343: gap of 100 bp
* 188344 194407: contig of 6064 bp in length..
FEATURES
Source
1. 194407
Location/Qualifiers

```

```

/misc_feature
1. .1755
/clone="RP11-73711"
/clone_lib="RP11-11 Human Male BAC"
/misc_feature
1856. .3429
/note="assembly_fragment"
/misc_feature
3530. .5346
/note="assembly_fragment"
/misc_feature
5447. .7338
/note="assembly_fragment"
/misc_feature
7439. .10297
/note="assembly_fragment"
/misc_feature
10398. .13590
/note="assembly_fragment"
/misc_feature
13691. .16293
/note="assembly_fragment"
/misc_feature
16394. .19065
/note="assembly_fragment"
/misc_feature
19166. .22713
/note="assembly_fragment"
/misc_feature
22814. .27165
/note="assembly_fragment"
/misc_feature
27266. .32377
/note="assembly_fragment"
/misc_feature
32478. .36572
/note="assembly_fragment"
/misc_feature
36673. .41657
/note="assembly_fragment"
/misc_feature
41758. .47024
/note="assembly_fragment"
/misc_feature
47125. .52938
/note="assembly_fragment"
/misc_feature
53039. .59129
/note="assembly_fragment"
/misc_feature
59230. .65896
/note="assembly_fragment"
/misc_feature
65997. .89427
/note="assembly_fragment"
/misc_feature
89538. .95785
/note="assembly_fragment"
/misc_feature
95886. .102762
/note="assembly_fragment"
/misc_feature
102863. .110899
/note="assembly_fragment"

```

```
misc_feature      111000. .121002 /note="assembly_fragment"  
misc_feature      121103. .133048 /note="assembly_fragment"  
misc_feature      133149. .145755 /note="assembly_fragment"  
misc_feature      145856. .165006 /note="assembly_fragment"  
misc_feature      165107. .188243 /note="assembly_fragment"  
misc_feature      188344. .194407 /note="assembly_fragment"  
clone_end:T7  
vector_side:right"  
BASE COUNT      50993 a 43965 c 46383 g 50458 t 2608 others  
ORIGIN
```

Query Match	17.6%;	Score 553.2;	DB 54;	Length 194407;
Best Local Similarity	90.4%;	Pred. No. 1e-117;		
Matches 665; Conservative	0;	Mismatches 48;	Indels 23;	Gaps 6

QY	1677	gctgcgctgaacatcctlaaagttgtctgtgaglttgaccacaagaatacagagatgcc	1736
Db	3297	gctgcactgaacacaccttaaggtgtctgtgagttggaccacaagaatgacagatgcc	3238
QY	1737	agaacaggaacgcgaaccaatgtctgtgtgtggaggtgtgaaccttttcggccatgaa	1796
Db	3237	agaacaggaatggatcatatgtctgtgtgtggaggtgtgaaccttttcggccatgaa	3178
QY	1797	ccattataaatcccaacatatatactgaaataactgtaaccttttgaanaatttgaat	1856
Db	3177	ccattataaatacccaacatatatactgaaataattttaaaactgctttgaaatatttgaat	3118
QY	1857	ttctgatacttccagltggtggtgagagacacgltggtlaaagatgtgtggtgagcagcag	1916
Db	3117	ttctgatacttccagltggtggtgagagacacgltggtlaaagatgtgtggtgagcagcag	3060
QY	1917	aagacacagaaacacacaagagagcggtctgtggtcc-ggctgacctgtgtggtttgttg	1975
Db	3059	aacacacacag-----ggaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	3007
QY	1976	tgtatgtccactcgtgtgacctgtggtgtccctacgcaatagcagctgtcgtgtgtggaaga	2035
Db	3006	tgtacgtccactcgtgtgacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2947
QY	2036	ggctgtccacagcagctgtgttctcccggtgacacacagcagatccacacctgtgtgtgtgt	2095
Db	2946	ggctgtccacagcagctgtgttctcccggtgacacacagcagatccacacctgtgtgtgtgt	2887
QY	2096	tgttgtgttttttttttccctgtg--tgaagaagaagaacggtgacagaccttctcaagc	2153
Db	2886	tgttgtgttttttttttccctgtg--tgaagaagaagaacggtgacagaccttctcaagc	2827
QY	2154	tgtgtcactcagacacatltgtggacaacacctgtgacagccatgcagagagagcccttga	2213
Db	2826	tgtgtcactcagacacatltgtggacaacacctgtgacagccatgcagagagagcccttga	2776
QY	2214	ccgtcccccagagctaaagcaccagagaaatacaaatgtcttctactcagcgtgaccaca	2273
Db	2775	ccagccccagagctaaagcaccagagaaatacaaatgtcttctactcagcgtgaccaca	2716
QY	2274	cttttctagtgtgccaaggt	2333
Db	2715	c--ttctagt	2658
QY	2334	tctcttccacagtgatctgtatcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2393
Db	2657	tctcttccacagtgatctgtatcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2598
QY	2394	tataataatttcat 2409	
Db	2597	tataataatttctct 2582	

Db 149889 TGCTGGATTACTGGCATGAGCCACCGTGCCGACGACACTATATAAAATTTTCATTTGA 149830
Oy 2414 gaatttcgaattgtatcgtgttaaalagcacagtttgaaaactgtctgagactgactt 2473
Db 149829 GAGTTTCTCAACTGTATCTAATTATATGCGACAGTTTGAAACTTTTCTGAGACTGACTT 149770
Oy 2474 tatcaataatctaaccgacaaagatcatatccatgtgtatgtgttagacattttatt 2533
Db 149769 TATCAATAAATCTAACCCGACAAAGATCATATCCAGCTGTATGATTAGACATTTTATT 149710
Oy 2534 catlgactaaccagagacagtttcagtgatgcaaatgtgtgccctctgtgttcagctgaa 2593
Db 149709 CATTGACTAACCCAGCACCATTTCAAGTATGCAAAATTAAGTGCCCTATGTTGACCTGAA 149650
Oy 2594 acagtcctggaacttccaacaaacttgaataagctctccacacagltgtataaattggacaat 2653
Db 149649 ACAGTCTCTGGACTTTCAAAAAACCCTGAATAAGTCTCCACACAGTTGTATTAATTGGACAAT 149590
Oy 2654 ttaggaaattttaaactttagatgatcattgtgttccatttttattcattttatttg 2713
Db 149589 TTAGGAATTTTAAACTTTAGATGATCATTTTGGCTTCCATTTTATTATTATTATTG 149530
Oy 2714 ttaatgcaaacagagactt---aaatgaacttgcctctgtttaaagatatlaaaaa 2769
Db 149529 TTTAATGCAAAACAGACTTTAAACAAATGAACACTTTGATCTCCATTTTAAATATTTCTTAAAGA 149470
Oy 2770 acatgtgtatctatacatatgctcttgagagacttaagcttccactacactacagatat 2829
Db 149469 AAATCATGTGCTATACATACAGGCTCTTGAGGACTTAGCTTTCACACTACACTACAGATAT 149410
Oy 2830 gatcccatgtagtcacatataaacctgcagagtgatgtttccagagtgtcgtactgttta 2889
Db 149409 GATCTCCACATAGTTCACATATAAACTGCAGACTGATTTTCCAGAGTGCCTGCATACTATTA 149350
Oy 2890 attacatctccattagggctgaaaagaatgacactagtttctgtatacagctgtgtgt 2949
Db 149349 ATTACATCTCGATTAGGGCTGAAGATGATGACCATTGTTCTTTATACAGTTGTGTGCT 149290
Oy 2950 tttagatgtgtgttactgtacacagaagtgtgtgcactgaagc 2992
Db 149289 TTTGATGTTGTGTACTATACACAGAAGTGTGTGCACTGGGCC 149247

RESULT 12
AC068845/c AC068845 194407 bp DNA HTG 04-SEP-2000
LOCUS Homo sapiens chromosome 19 clone RP11-73711 map 19, WORKING DRAFT
DEFINITION
SEQUENCE, 27 unordered pieces.
ACCESSION AC068845
VERSION AC068845.3 GI:9966312
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194407)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 19, clone RP11-73711
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194407)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczký,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidrim,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,

TITLE
JOURNAL
COMMENT

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Tajamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:8247880.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7549
Center clone name: 737_I_1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181875 bases at least Q40
Consensus quality: 187842 bases at least Q30
Consensus quality: 190220 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 191807; sum-of-ctrls
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1755: contig of 1755 bp in length
* 1756 1855: gap of 100 bp
* 1856 3429: contig of 1574 bp in length
* 3430 3529: gap of 100 bp
* 3530 5346: contig of 1817 bp in length
* 5347 5446: gap of 100 bp
* 5447 7338: contig of 1892 bp in length
* 7339 7438: gap of 100 bp
* 7439 10297: contig of 2859 bp in length
* 10298 10397: gap of 100 bp
* 10398 13590: contig of 3193 bp in length
* 13591 13690: gap of 100 bp
* 13691 16293: contig of 2603 bp in length
* 16294 16393: gap of 100 bp
* 16394 19065: contig of 2672 bp in length
* 19066 19165: gap of 100 bp
* 19166 22713: contig of 3548 bp in length
* 22714 22813: gap of 100 bp
* 22814 27165: contig of 4352 bp in length
* 27166 27265: gap of 100 bp
* 27266 32377: contig of 5112 bp in length
* 32378 32477: gap of 100 bp
* 32478 36572: contig of 4095 bp in length
* 36573 36672: gap of 100 bp
* 36673 41657: contig of 4985 bp in length
* 41658 41757: gap of 100 bp
* 41758 47024: contig of 5267 bp in length
* 47025 47124: gap of 100 bp
* 47125 52938: contig of 5814 bp in length
* 52939 53038: gap of 100 bp

```

AC011492/c
LOCUS          AC011492    157803 bp    DNA             HTG              04-MAY-2000
DEFINITION     Homo sapiens chromosome 19 clone CTB-187L3, WORKING DRAFT SEQUENCE,
ACCESSION      AC011492
VERSION        AC011492.5  GI:7690131
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          1 (bases 1 to 157803)
JOURNAL        DOE Joint Genome Institute.
REFERENCE      Sequencing of Human Chromosome 19
AUTHORS        Unpublished
TITLE          2 (bases 1 to 157803)
JOURNAL        DOE Joint Genome Institute.
REFERENCE      Direct Submission
AUTHORS        Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
TITLE          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL        On May 4, 2000 this sequence version replaced gi:7458736.
COMMENT        -----Genome Center
                Center: Joint Genome Institute
                Center Code: JGI
                Web site: http://www.jgi.doe.gov
                -----
                Project Information
                Center Project Name: 136279, BC101302
                Center clone name: CIT978SKB_187L3
                -----
                Summary Statistics
                Consensus quality: 152548 bases at least Q40
                Consensus quality: 153824 bases at least Q30
                Consensus quality: 154424 bases at least Q20
                Estimated insert size: 156750; agarose-fp estimation
                Estimated insert size: 157103; sum-of-contigs estimation
                Quality coverage: 9.52 in Q20 bases; agarose-fp estimation
                Quality coverage: 9.5 in Q20 bases; sum-of-contigs estimation.
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 8 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                1
                *      2374: contig of 2374 bp in length
                *      2375      2474: gap of unknown length
                *      2475      4521: contig of 2047 bp in length
                *      4522      4621: gap of unknown length
                *      4622      7566: contig of 2945 bp in length
                *      7567      7666: gap of unknown length
                *      7667      12083: contig of 4417 bp in length
                *      12084      12183: gap of unknown length
                *      12184      18288: contig of 6105 bp in length
                *      18289      18388: gap of unknown length
                *      18389      63650: contig of 45262 bp in length
                *      63651      63750: gap of unknown length
                *      63751      111373: contig of 47623 bp in length
                *      111374      111473: gap of unknown length
                *      111474      157803: contig of 46330 bp in length.
                *
                Location/Qualifiers
                1. 157803
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="19"
                   /clone_lib="Caltech human BAC library B"
                   /clone_lib="38390 c 39640 g 38079 t 711 others
BASE COUNT     40983 a 38390 c 39640 g 38079 t 711 others
ORIGIN
Query Match    22.4%; Score 705; DB 35; Length 157803;

```

[illegible]

[/db_xref=GI:9754866](#) "translation="MYKVPDPSRMOSTYVSYGMRGAVPRRYEYFPEVPLLVOVELS
VGGQOFGKGMKRPVKVHTDARLTLOSEPRLPERLYVNGRESEENYNKSEISOVF
EIALKRLNLPVNESEFPLTOVARESGPHMKNEFTRVSVGEFVGESEKSKISKNNAA
RAVEQLRLPLPAVERVKPRIKKSQPTCKLOTAPDYGGQMPISRAOIQAKKE
KEPEYMLTERGLPRRREFVMQVKGHTHAGAGTNKKVAKRNAENMLEILGFVPQ
AQPAPALKSEETPVKKGPDGRKTFEFPSPGDENGTSNKEDEFRMYLSHQLPAG
ILEMPEVAQAQVTSVSGGHTTKDFRAGPFAKATVTAMIARELLYGGTSPTAETILKS
NISGHPHGPRTRPSQQLYYLSRQAGFQVEYKDFPKNNKNECVSLINCSSQPRLYVSH
GICKDVESCHDMAALNLIKLSLELDQOSTEMPRTGNGPVASACGR"

```
polyA_signal      3043.  .3048
```

```
/gene="Stau"
```

BASE COUNT	817 a	778 c	749 g	713 t	2 others
ORIGIN					

Query Match	42.28;	Score 1326.6;	DB 74;	Length 3059;
Best Local Similarity	85.38;	Pred. No. 1.2e-297;		
Matches 1548; Conservative	0;	Mismatches 239;	Indels 28;	Gaps 5;

QY	89	ccctcccccccgccggccggcgccgcctctctccacggccactccgcctctctcc	148
Db	13	CTTCTGACCCCGCGCCCGCGCGCGCGCTTCTCCGACAGCCACTCCGTCCTTCCCTCC	72
QY	149	ctctgcctctctctcctcctcttttctctctctctctccctcctcgcgcgcacgc	208
Db	73	CCCGGCTCTTCTTCTCCTCGCCCTCTTCTTTGGCTT-----TTCTTCTCGCGCGCGCC	125
QY	209	cagcaccgcccggccggggagcagctc-ggagcagcagccagaanaagcataccctactg	267
Db	126	CAATACCGCGCGCCCGCGGAGCAGCTCTGGGGAAGCAGCCAGAAGTATAGCTTCTACCA	185
QY	268	tagaactaaatgcactgtgtgcatgaaacttggaaaaaaccaatgtataagcctgtgacc	327
Db	186	TTGAGCTCAATGCACCTGTCGCTGAACACTGGAAGAAACCAATGTATAGCCCGCTGACCC	245
QY	328	cttactctcgatgcagtcaccactaactacaatgagagaggtgtctatccccga	387
Db	246	CCCACTCTCGGATGCAGTCCACCTACAGCTATGGCATGGCGTGAGGTGCTATCCCCCA	305
QY	388	ggtacttttaaccatttccagttccaccttacttataaagtggaacttctctggag	447
Db	306	GATACTTTTACCCATTTCAGTTCACCGCTTACTCTACCAAGTTGAGCTCTCCGCGGAG	365
QY	448	gacagcaatttaatgycgaaggaagaacaagacagcgtctcgaaacacagatgtctgtcca	507
Db	366	GACAGCAGTTTAAATGGGAAGGGAACATGAGACCACCTGTGAACAATGATGCCACTGCC	425
QY	508	aagcgttgagatccctgcagaatlaagccccctgccagagaggtgtgagtgtaatgaagag	567
Db	426	GAGCCCTGAGSACTCTGCAGAGTGAGCCCCCTACCAGAGAGGCTTAGAGTTAATGGAAGAG	485
QY	568	aatccgaagagaanaaaatctcaataaactgaaataagtcagtglttgagatgcacta	627
Db	486	AATCAGAAGAGAAAAACCTTAATTAATCAGAAATTAAGCCAACTGTTGAGATTGCCCTGA	545
QY	628	aacggaactgtcctgtgaatttcg-----agtgccccggagagtg	669
Db	546	AGCGGAATTTGCCCTGTGATTTTGAGTCTTTCCTCTGACACAGGTGGCCCCGGAGAGTG	605
QY	670	gccccacccacatgaagaactctgtgaccaaggttcggttggggagttgttgggggaag	729
Db	606	GCCCCCACACATGAAGAACTTTGTGACCAGGGTTTCAGTTGGGGAGTTTGTAGGGGAAG	665
QY	730	gtgaagggaaaagcaagaagatttcaaagaaaaatgcccacatagctgttcttgaggagc	789
Db	666	GAGAAGGGAAGCAAGAAGATCTCAAGAAGAATGGCGCAAGGGCCGTGCTGAGCAGC	725
QY	790	tgaagaagttaacgccccctgcctgcagttgaaagagtaaaagcctagaaatcaaaaaaanaa	849
Db	726	TAAGGAGGCTGCCACCTTACCTGCTGTGGAGCGAGTGAAGCCCAAGATCAAGAGAAAA	785
QY	850	caaaaccatagtcaagccacagacaagcccaagaatataatgccaagggatcaatccgatta	909

Db	786	gTTCAGCCACCTGCAAGCTACAGACTGCCCCCGACTATGGCCAAAGGATGAATCCTATT	845
QY	910	gccgactggtccagatccagcagcgaanaaaggagaagagccagagtacaagctcctca	969
Db	846	GCAGACTGGCACAGATCCAGCAGCAGCAAAAAGAGAGAAGAGAGCCGGAGTACATGCTCCTTA	905
QY	970	cagagcagggcctcccgcgccgcagggagtttgtatgcaggtgaagtttgaaaccaca	1029
Db	906	CAGAACGAGGCTCTCCACGTCGACGGAGTTGTGATGCAGGTGAAGTTGGGCATCACA	965
QY	1030	ctgcagaaggaacggygcaccaacaagaagtylgccaagcgcaatgcagccgagaacatgc	1089
Db	966	CCGCAAGAGGAGCAGGTACCAATPAGAAGGTGGCCAAGCGCAATGCTGTGAGAACATGC	1025
QY	1090	tygagatccttggtttcaagtlcccgagcgcgagcccaaccaaccgcaactcaagtca	1149
Db	1026	TGGAGATCCTGGGGTTCAAAGTCCCCAGGCGCAGCCTGCCAAGCCAGCACTCAATCAG	1085
QY	1150	aggagaagacaccataaagaaccaggggatygaaagaaagttaacctttltgaaactg	1209
Db	1086	AAGAGAAGACTCCAGTAAGAAGAACCGAGAGATGGAAGGAAGTAACTTTTGAACCTA	1145
QY	1210	gctctgggagtgaaatbgygactagttaataagagatgagtlcgaagtgcctatac	1269
Db	1146	GCCCTGGGGATGAAATGGAACCAAGTAATPAAAGAGGATGAGTTCAGATGCCCTATCTTA	1205
QY	1270	gtcatcagcagctgcctgtctggaattctcccatgtgtcccgaggtgcgccaggtgtag	1329
Db	1206	GCCATCAGCAGCTGCCAGCGGGGATTTCTCCCATGTGTGCAGAGGTTGCCCAGGCTGTGC	1265
QY	1330	gagttagtcaagacatcacaccaagaatttlaaccagggcagctccgaatcctctgccaag	1389
Db	1266	GGGTAGTCAAGGACACCATACCAAGATTTCCGCCAGGGCAGCTCCAATCCTGCCAAGG	1325
QY	1390	ccaacggttaactgcatgatagcccgaagagtttgtatggyggcacctgcgccacagcg	1449
Db	1326	CCACGGTAAGTCCATGATAGCCCGAGAATTGTGTACGGGGGCACCTCGCCACACAGCCG	1385
QY	1450	agaccatlltaagaataacatctctcaaggccaagtaaccccaatgaacctctcaagagac	1509
Db	1386	AGACCATTTTAAGAGTAACATCTTTCAGGACATGTACCCCATGACCTCGCACAAAGAC	1445
QY	1510	ccctcgaagcaactgactatcttccagagtlccagagattccaggttgaaatacaaaagact	1569
Db	1446	CCTCTGAGCAACTGTACTACCTTCCAGAGCCCGAGGATTCAGGTGGAATACAAAGACT	1505
QY	1570	tcoccaaaaacaacaagaacgaatttgtatctcttatcaatitgtctcctcagccaactc	1629
Db	1506	TCCCCAAGAACAACAAGACGAGTGTATCTCTCATCAACTGCTCTTCACAGCCACCTC	1565
QY	1630	tgatcagccatggtatcgcgcaagagatgtgaggtcctcgtccatgatatggtcgtcgtaaca	1689
Db	1566	TCGTCAGCCATGGCATCGGCAAGGATGTGGAGTCCGTGCATGATATGCTCGCACTGAACA	1625
QY	1690	tcttaaaagttcgtctgtagttggaaccaacaagaatcacagagatgccaagaacaggaacg	1749
Db	1626	TCTTAAAGCTGCTGTCTGAGTTGGACCAACAGAGACAGAGATGCCAAGGACAGGAAACG	1685
QY	1750	gaccaatgtctgtgtgtggyaggtgctgaacctttctcgtgccatgaacattataaaatc	1809
Db	1686	GACCAGTTTCAGCGGTGCGGAGGTGCTGAACCTATTCTGGCCATGAACCATTTATAAAA-C	1744
QY	1810	ccaacatatatactgaanaatactga-aactgcttgaanaatttggaaatttctgatacctc	1868
Db	1745	CCAACATATATATACCGAAAATACTGAGAACTGCTTTGAATAATTGGAAATTCTGATAACTC	1804
QY	1869	caatgggcccagagaga 1883	
Db	1805	CAGTGGGCCAAGACA 1819	

RESULT 11

QY	328	cttactctcgatgctcagttccacctataactacaacatlgagagagtgctctatccccoga	387
Db	246	ccccactctcggatgacgtccacctacagactatgacatgctggaggtgcctatcccccca	305
QY	388	ggtacttltaccatctccagltccaccttacttatactcaagtgaacttctgtggag	447
Db	306	gatacttattaccatattccagttccaccggttacttaccaaagttgagctctccgtggag	365
QY	448	gacagcaatlttaatgycaaaagaaagacacagacagagctgcgaacacgatactgtgcc	507
Db	366	gacagcagctttaatgggaagggaagatgagaccacactgtgaacatgatgccactgcc	425
QY	508	aagcgtltgagatccttcagaaatgagccccctgcagagagaggtcgtgaggtgaatgaag	567
Db	426	gagccctgagactcttcgacagagtgaaccccttaccagagagcctagaggttaattgaag	485
QY	568	aatccgaagaagaanaatctcaataaatctgaanaagtcaagtgtltgagattgcactta	627
Db	486	aatcagaagaagaaccttaataatcagaataaagccaagtggttgagattgcgctga	545
QY	628	aacggaactgctgtgaatttcgaggtgycgcggagaggtgycaccaaccacatgaaga	687
Db	546	agcggatattgctgtgaattttgaggtgccccggagagagtgccccacacatgaaga	605
QY	688	acttgtgaccaaagttctcgtlttgaggagtttctgtgaggaaagtgaagggaanaagcaaga	747
Db	606	acttgtgaccacagggtttcagttgggagttgttaggggaaggagaagggaanaagcaaga	665
QY	748	agatttcaaaaaaatgycgcgcatagtctgttcttgaaggagctgaagaagttaccgccc	807
Db	666	agatctccaagaagaatgccccgaaggccgctgttgagacagcttaaggaggctgccacctc	725
QY	808	tgcttcgagttgaacgagtaaagcctagaatcaaaaaaagaaaacacacatagtcaagc	867
Db	726	tacctgctgtgagcgagctgaagccacagaatcaagaagaaaagtcagcccactgcacaagc	785
QY	868	cacagacaagcccagaatatgyccagggagatccaatccgatltagccgactgccccagatcc	927
Db	786	tacagactgcccccgactatgccccaaaggatgaatcctattagcagactggcacagatcc	845
QY	928	agcagcgcaaaaaagagaagagagccagagtacacgctctctcacagagcgaagcctccgc	987
Db	846	agcagcgcaaaaaagagaagagagcccgagtagtactcttcaagaacgaggtctttccac	905
QY	988	gcccgaaggagttgtbatagcaagtgaaaggttgagaaacacacactgcagaagaacgycga	1047
Db	906	gttcgacaggagttgtgatgcagggtgaaggttgccatcacaccgacagagagcaggtga	965
QY	1048	ccaacaagaagtgyccaagcgcacatgcagccgagacacatgctgagatccttgtttca	1107
Db	966	ccaataaagaagtgcccaagccgaatgctgtgagaaacatgctggagatcctggggttca	1025
QY	1108	aagtcccgcgagcgcgaccccaacacccgcacactcaagtacagagagaagacacccataa	1167
Db	1026	aagtccccccgagcgcgacccctcccaagccacgacactcaaatcagaagagaagactccagtaa	1085
QY	1168	agaaaaccagggagatggaagaaaaglaaccttlttgaacctgycctcgtggagatgaataatg	1227
Db	1086	agaaaccagagatggaagaaagtaacattttttgaaccttagccctgggagtgaaaaatg	1145
QY	1228	ggactagtaataaagagatgagltcagagatgcttatactaatgaatcacagcagctgcctg	1287
Db	1146	gaaccagtaataaagagagatgagttcagagatgccccctattcttagccatcagacagctgccag	1205
QY	1288	ctggaattcttcccatggtgccccgaggtgcgccagagctgtagagttagtcaagaacatc	1347
Db	1206	ccggagattctccccatggtgccagagaggttgccccagcgctgtcggggttagtcaagacaccc	1265
QY	1348	acaccaaaagatttaccagagcagctccgaatcctgcgcaagcgccacggttaactgcatga	1407
Db	1266	ataccaagaatttgcgcagcgacgctccaataatcctgcgaagggccacggtaactgccatga	1325

QY	1408	tagcccgagagtgtgltgtlatatggggaccctgcgccacagccgagaccattttaagaata	1467
Db	1326	TAGCCCGAGAATGTTGTACGGGGGCCACTCGCCACAGCCGAGACCATTTTAAAGAGTA	1385
QY	1468	acatctctcaggccaacglaccocatygacctctcacgagaccctctgagcaactygaact	1527
Db	1386	ACATCTCTTCAGGACACATGTACCCCATGAGCCTGCACCAAGACCCTTGAGCAACTGTACT	1445
QY	1528	atcttccagagtcaccaaggtlccagtgttacaagaactccccaaaacaacaaga	1587
Db	1446	ACCTTTCAGAGCCACGGGATTCAGGTTGAATACAAGACATTCCCCAAGAACAACAAGA	1505
QY	1588	acgaatttgtatctcttatcatattgtctcctctcagccacctctgatcagccatgyratcg	1647
Db	1506	ACGAGTGTGTATCTCTCATCAACTGCTCTTCACAGCCACCCTCTGTCAGCCATGGCAATCG	1565
QY	1648	gcaaagatatgtgagtcctcgccatgatatgytgcgtcgtaaacatcttaaagtlgtcrtcy	1707
Db	1566	GCAAAGATGTGGAGTCTCTGTCATGATATGGCTGCACCTGAACATCTTAAAGCTGCTCTCG	1625
QY	1708	agtttgaccacaacaagtacagagatgccaaagaacagyaacagyaaccaatgtctgtgtg	1767
Db	1626	AGTTGGACCAACACAGACACAGAGATGCCAAGACAGACAGGAALCGGACCAGTTTCAGCGGTGCG	1685
QY	1768	ggaggtgtcgaaccttttctgycgatgaaccaattataaaatcccacaatatatactyaaa	1827
Db	1686	GGAGGTGCTGAACCTATTCTGCGCATGAACTATTATAAAA-CCCACATATATATCCGAAA	1744
QY	1828	atactga-aactgcttggaaaaatttggaaattctgatacctccagtyggccgagaga	1883
Db	1745	ATACTGAGAACTGCTTTGAAAAATTGGAAATTTCTGATAACTCCAGTGGGCCAACAGACA	1801

```

RESULT 10
AF227200 LOCUS AF227200 3059 bp mRNA ROD 09-AUG-2000
DEFINITION Rattus norvegicus staufen isoform Stau+16 (Stau) mRNA, complete
cds, alternatively spliced.
ACCESSION AF227200
VERSION AF227200.1 GI:9754865
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3059)
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M.,
Desgroselliers, L., Kuhl, D., Richter, D. and Kindler, S.
TITLE Two rat brain staufen isoforms differentially bind RNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3059)
AUTHORS Monshausen, M. and Kindler, S.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) University of Hamburg, Institute for Cell
Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg
20246, Germany
FEATURES
source
1.3059
location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1.3059
/gene="Stau"
227.1714
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila staufen;
contains four RNA-binding domains; rStau+16"
/codon_start=1
/product="staufen isoform Stau+16"
/protein_id="AAF98119.1"

```

Db	855	TCAGAAGAAAAGTCAGCCACCCTGCA-----AGACAGCCCCGGATTATGGCCAAGGGA	908
QY	898	tcaatccgattagccgactgccccgatccagcagcagcaaaaaaggagaagcaccagt	957
Db	909	TGAATCCATTAGTAGACTTGCACACATCCAGCAGGCAAAAAAGAGAAGAGCCACAGT	968
QY	958	acacgctccctcacagagcgagcgctcccgccgccgagggagttgtatgcaggtaaag	1017
Db	969	ACATGCTCCTTACAGAACGAGGCTCTTCCACGCTCGCAGGGAGTTGTGATGCAGGTAAGG	1028
QY	1018	ttgnaaaccaactgcagaaggaagcgccaccaacaagaagtgccaaagcgcaatgcag	1077
Db	1029	TTGGGCAATCAGACTGCAGAAGAGTGGGTACCAATAAGAAGGTGGCCAAGCGTAATGCTG	1088
QY	1078	ccgagaacatgctgagatcccttgtttcaaaagtcgccgagcgccaccacaaccgcg	1137
Db	1089	CTGAGAACAATGCTGGAGATCCTGGGGTCAAGTTCCCCAGGGCAGCCTGCCAAGCCAG	1148
QY	1138	cactcaagtcagaaggagaagacacccataaagaacaacgaaggatggaagaaagtaact	1197
Db	1149	CACTCAAAATCAGAAAGAGAAGACTCCAGTAAAGAAACCAGGAGACGGAAGAAAGTACGT	1208
QY	1198	ttttgaacctgctctggygatgaaaaatggygactagtaataaagaaggaatgagtcaga	1257
Db	1209	TTTTGAACCTAGCCCTGGGATGAATAAGAACTAGTAACAAGAGACGAGAGTTCAAGA	1268
QY	1258	tgcctatctaagtcacagcagctgctgtgnaattcttcccatggtgccgaggtcg	1317
Db	1269	TGCTTATCTTAGCCATCAGCAGCTGCCAGCTGGAATTCCTCCCATGCTGCCGGAAGTTG	1328
QY	1318	cccagctgtgaggagttagtcaaggaatcacaccaagaattttaccagggcagctccga	1377
Db	1329	CCGAGGCTGTGGGGGTAGTCAAGGACACACCAAGATTTCACCAGGGCAGCTCCAA	1388
QY	1378	atcctgccaaagcgcaagtgtaactgcatgtagcccgagagttgttatggyggcaact	1437
Db	1389	ATCCTGCCAAGGCAACGGTAACTGCCATGATAGCCCCGAGAGTTGTTACGGGGGCACCT	1448
QY	1438	cgccacagccgagaccattttaagaaataacatctcttcagggccaagtlaccccatgac	1497
Db	1449	CGCCACAGCCGAGACCATTTTAAAGAGTAACATCTCTCAGGCCACGTAACCCCATGGAC	1508
QY	1498	ctctcagagaccctctgagcaactgactatcttccagagtcgaagatccaggttcg	1557
Db	1509	CTGCACCTAGACCTCTGAGCAACTGTACTACCTTCCAGAGCCACGAGATTCAGGTTG	1568
QY	1558	aatacaaaagacttccccaaaaaacaacaagaagcaatttgtatcttataaattgtcct	1617
Db	1569	AATACAAAGATTTTCCCAAGAACACAGAAGAGAGTGTGTATCTCTCATCACTGCTCCT	1628
QY	1618	ctcagccacctctgatcagccatggtatcggcaagagatgtgagtcctcgcatagtatgg	1677
Db	1629	CACAGCCGCTCTCTGTCAGTCAATGGCATCGGCAAGAGATGTGAGTCCCTGTCATGATATGG	1688
QY	1678	ctgcgtcgaacatcttaaaagtgtcgtctgagttggaaccaaaagtlacagagatgccaa	1737
Db	1689	CTGCACCTGAACATTTTAAAGCTGCTGTGAGTTGGACCAACAGACAGAGATGCCAA	1748
QY	1738	gaacaggaacggaaccaatgtctgtgtgtggyaggtgtgtaacctttctgccaatgaac	1797
Db	1749	GAACAGGAATGAGACCAAGTTTCAGCGGTGGGGAGGTGCTGAACCTTTTCTGGCCACAAC	1808
QY	1798	cattataaaatcccaacatatatactgaaaaatactga-aactgctttgaaaatttggaa	1856
Db	1809	CATTATATAAA-CCCAACATATATACTGAATAATACTGAGAACTGCTTTGAAAATTTGGAA	1867
QY	1857	ttctgataccttcagtggtggccgagaga	1883
Db	1868	ATCTGATACTCCAGTGGGCCAAGACA	1894

RESULT 9
AF290989

LOCUS	AF290989	3041 bp	mRNA	ROD	09-AUG-2000
DEFINITION	Rattus norvegicus staufen isoform Stau-I6 (Stau) mRNA, complete cds, alternatively spliced.				
ACCESSION	AF290989				
VERSION	AF290989.1 GI:9755162				
KEYWORDS					
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	1 (bases 1 to 3041) Monshausen,M., Putz,U., Rehbein,M., Schweizer,M., DesGroselliers,L., Kuhl,D., Richter,D. and Kindler,S. Two rat brain staufen isoforms differentially bind RNA				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 3041) Monshausen,M. and Kindler,S.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (28-JUL-2000) Institute for Cell Biochemistry and Clinical Neurobiology, University of Hamburg, University Hospital Eppendorf, Sudefeldstr. 24, Hamburg D-22529, Germany				
AUTHORS	Location/Qualifiers				
JOURNAL	1. 3041				
FEATURES	/organism="Rattus norvegicus"				
source	/strain="Sprague-Dawley"				
	/db_xref="taxon:10116"				
	/tissue_type="brain"				
	/dev_stage="adult"				
	1. 3041				
gene	/gene="Stau"				
	227. 1696				
CDS	/gene="Stau"				
	/note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau-I6"				
	/codon_start=1				
	/product="staufen isoform Stau-I6"				
	/protein_id="AAF98177.1"				
	/db_xref="GI:9755163"				
	/translation="MYKPVDPHSRMOSTYSYGMKRGAYPRRYFPVPVPLLYQVELS VGGQFNGKMRPVPKHDATAARLRLTQSEPLERLEVNGRESEENLNKSEISOVE EIALKRLNLPVNEFVARESGPRPHMKNFVTRVSVEFVGESEKSKISKNAARAVLEQ LRLPLPAVERVKPRIKKKSQPTCKLQTAADYGQGMNPISRLAQIOAKKEPEYML LTERGLPRRREFVMQYKVGHHTAEGAGTNKKVAKRNAEMLIEIGKVPQAQPAK ALKSEKTPVKKPGDKRKVTFFESPGEENGSNKEDEFMRPVLSHOQLPAGILPMVP EVAQAVGSQGHHTKDFARAAPNPAKATVTAMIARELLYGTSPTAETILKSNISSGH VPHGPRTRPSEQLYLSRAQGEVEXKDFPKNNKNECVSLINCSQPLVSHGIGKDV ESCHDMAALNILKLSELDDQSTEMPRTGNGPVSA CGRC"				
	3025. 3030				
polyA_signal	/gene="Stau"				
	3025. 3030				
BASE COUNT	814 a	771 c	747 g	707 t	2 others
ORIGIN					
Query Match	43.1%; Score 1354.6; DB 74; Length 3041;				
Best Local Similarity	86.1%; Pred. No. 0;				
Matches 1548; Conservative	0; Mismatches 239; Indels 10; Gaps 4;				
QY	89	ctcccccccgccgccccgccccgccccctcccaagcgaactcgcctctccctcc	148		
Db	13	CTTCTGACCCCGCGCCCGCGCGCCCTTCTCCGACGCACTCCGTCTTCCCTCC	72		
QY	149	cttcgctccctctctcctctcccttttctcctctctccttccctcctcgcgcgaacg	208		
Db	73	CCCGGTCTTCTTCTCTCGCCCTCTTCTTTTGCTT-----TTCTTCTCTCGCGCGCC	125		
QY	209	caggaccgccccgccccgccccgccccgccccgccccgccccgccccgccccg	267		
Db	126	CAATACCGCCCGCCCGGAGAGCTCTGGGGAGAGCAGCCAGAAAGTATAGCTTCTACCA	185		
QY	268	tagaactaaatgcaactgtgcatgaaacttgaaaaaacaatgtataagcctgttgacc	327		
Db	186	TTGAGCTCAATGCACCTGTGCGTGAAGACTGAAAGAAACCAATGTATAGCCCGTGACC	245		

[illegible][illegible]

Db 2577 GATCTCCATGTAGTCCATATAAACCTGCAGAGTGATTTTCCAGAGTCTCGATACTGTTA 2636

QY 2890 attacatctccatcaggcgtgaaagaatgaacctacgtttctgtatcacagctgtgttgc 2949
|||||

Db 2637 ATTACATCTCCATTAAGGCGCTGAAAGAATGACCTACGTTTCTGTATACAGCTGTGTGCT 2696

QY 2950 ttgatgtgtgttaactgtacacagaagtgtgtgcaactgagcctgtcggtgtgtccgta 3009
|||||

Db 2697 TTTGATGTGTGTTACTGTACACAGAAGTGTGTGCACTGAGGCTGTGCGTGTGCCGTA 2756

QY 3010 tggaaaacctggtagccctgcgagtaagtactgtctccatcattcatttaccgtgaat 3069
|||||

Db 2757 TGGAAAGCCTGGTAGCCCTGCGAGTTAAGTACTGCTTCATTCATTTGTACGCTGAAT 2816

QY 3070 ttttccccatggaatgtaaagtaaaactaagtgttgtcatcaataaatgtaatact 3129
|||||

Db 2817 TTTTCTCCCCATGGAATGTAAGTAAACCTTAAGTGTGTTCATCAATAATGTAATACT 2876

QY 3130 aaa 3132
|||

Db 2877 AAA 2879

RESULT 7
AL133174/c 113836 bp DNA PRI 14-JUN-2000
LOCUS Human DNA sequence from clone RP3-470L14 on chromosome 20, complete
DEFINITION sequence.
ACCESSION AL133174
VERSION AL133174.15 GI:8573761
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 113836)
AUTHORS Ramsay,H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Jun 20, 2000 this sequence version replaced gi:8388434.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP3-470L14 is from the library RPCI-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP3-470L14 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP3-470L14 is at 113836 in this
sequence. The true right end of clone RP1-155G6 is at 100 in this
sequence.

FEATURES
Source
Location/Qualifiers
1..113836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP3-470L14"
/clone_1id="RPCI-3"
3061..3077
/note="Weak data"
misc_feature
95583..95627
/note="Single clone region. pUC subcloned to give 12x
coverage. Assembly confirmed by restriction digest data."
BASE COUNT 33974 a 25115 c 24233 g 30514 t
ORIGIN
Query Match 43.2%; Score 1357.4; DB 34; Length 113836;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1763 gtgtggaggtgtcgtgaaccttttctgtgccaatgaaccattataaatacccaatatatac 1822
Db 34965 GTGTGGAGAGTGTCTGAACCTTTTCTGGCCATGAACCATTTATAAATCCCACATATATAC 34906
QY 1823 tggaaatactgaaactgtcttgaaaatttgaatttctgataccctcagttggtggcagag 1882
Db 34905 TGAATAATCTGAACCTGCTTTGAAAAATTGGAATTCTGTATACCTCCAGTGGCGGAGAG 34846
QY 1883 acaagtggtgtaaaagatltggtgcagcagcaggggaagacaacagaaacacaagagcg 1942
Db 34845 ACACGCTGGTAAAGATGTGGGCAGCAGCAGGGAAGACAACAGAAACACAAGAGGCGG 34786
QY 1943 ctgtggcc-ggctggaactgtgtcgtgggttgttgtatggtgccactcgtgtgacctggt 2001
Db 34785 CTGTGGCCGGCTGTGACTGTGCTGGGGTTGTGTGATGGCCACTCGGTGACCTGGCGGT 34726
QY 2002 ccctacgcaatagcagctgcctgtgtggggaagaaggctgccagccaagctgtgtcccg 2061
Db 34725 CCCTACGCAATAGCAGCTGCTGTGGGGAAGAAGGGCTGCCAGCCAGCTGTTCTCCCG 34666
QY 2062 ggaacacagcagatccacacacctgtggacccctcgttgttgttctttttccctgtgt 2121
Db 34665 GGACACAGCAGCAGATCCACACCCCTGGGACACTCCGTGTGTGTTTGTCTTTTCCCTGTGT 34606
QY 2122 gaaagaagaacgcaagcagccctctcaagctgtgtcactcagacacatgtggacaac 2181
Db 34605 GAAAGAAGAAACGCGACGACCCCTTCTCAAGCTGGCTCAGTCACTCAGACACATTGGGACAAAC 34546
QY 2182 cctgacagccatgtccagagagagagcctttgaaccggccccagagctaaagcaccagaga 2241
Db 34545 CCTGACAGCCATGCGACAGAGAGAGGCGCTTGAACGGCCCCAGAGCTAAAGACACAGAGA 34486
QY 2242 aaatcaaatgtctcctactcagcgtgacccaacttttctaagtgtgccagcgccccaccac 2301
Db 34485 AAATCAATGCTTCTACTCTCAGCGTGACCCAATTCTTAGTGTGCCACGGCCCCACACAC 34426
QY 2302 ctctcgagtaaccacacacatcaccactgtcttctctccaacagtgatctgtatctcta 2361
Db 34425 CTCCTGAGTACCAACACCATCACCACCTGCTTCTCTCAACAGATGATCTGTATCTTA 34366
QY 2362 gttcatatttctcttttgatgatgatgataataaataatttcatttgaatttct 2421
Db 34365 GTTTCATTATTTCTTTTGATGTGATATGACACTATATAAAATTTCATTGAGAAATTCT 34306
QY 2422 caattgtatctagtaataatagcacaggttggaaactgtctgagactgactttatcaata 2481
Db 34305 CAATTGTATCTAGTTAATAGCACAGTTTGGAAACTTGTCTGAGACTGACTTTATCAATA 34246
QY 2482 atctaaccgacaagaatcatatcatgtgtatgtgtgtagacatttcatcatgaact 2541
Db 34245 ATCTAACCGACAAGATCATATCCATGTGTATGTGTTAGACATTTTATTTCATTTGACT 34186
QY 2542 aaccagacagtttcagtgatgcaaatgtgtgccctctgtgttcagctgaacagtcct 2601

QY 2227 taaagcaccagagaaaatcaaatgctctcctactcagcgtgacccaactttctagtgtg 2286
|||||
Db 2591 TAAAGCACACAGAGAAAATCAAAATGCTTCCCTACTCAGCGTGACCCAACTTTCTAGTGTG 2650
QY 2287 ccacgccccacacccctcctcgtcagtaccccacacacatcacaccactgtcttctctccaacag 2346
|||||
Db 2651 CCACGGCCCAACACCTCCTGCGAGTACCACACACATCACCACACTTCTCTCCACACAG 2710
QY 2347 tgatctgtatctcttagttcattatcttcttcttcttgattgatatgaactatataaaatltt 2406
|||||
Db 2711 TGATCTGTATCTTAGTTTCATTATTTCTTTGATTGATATGACACTATATAAAATTTT 2770
QY 2407 cattgagaattctcaattgtatctagttaaatagcacagtttgaaaactgtctaga 2466
|||||
Db 2771 CATTGAGAAATTTCTCAATGTATCTAGTTAAATAGCACAGTTTGGAAACTTGTCTGAGA 2830
QY 2467 ctgacttatcaataatctaaccgacaaagatcatatccatgtgtatgtgtagaacatt 2526
|||||
Db 2831 CTGACTTTATCAATTAATCTAACCAGACAAGATCATATCCATGTGTATGTGTAGACATT 2890
QY 2527 ttatctcatgactaaccagagacagtttcagtgatgcaaatgtgtgccctctgttc 2586
|||||
Db 2891 TTTATTTTCATTCGACTAACCCAGACAGTTCAGTGATGCAAAATGTGTGCCCTCTGGTTC 2950
QY 2587 agctgaacacgtccctgagactttccaanaaccttgaaatagtcctccacagttgtataaatt 2646
|||||
Db 2951 AGCTGAACACAGTCTGGACTTTCAAAACCTTGAATAGTCTCCACACAGTGTATATAATT 3010
QY 2647 ggacaatttggaaattttaaacttctgatatcattgtgtccatttcttattcattttt 2706
|||||
Db 3011 GGACAATTTAGGAATTTTAAACCTTTAGATGATCATTTGGTTCATTTTTATTTCAATTTT 3070
QY 2707 attttgttaatgcaaacagagacttaaatgaacttgaatctctgtttaaagattataa 2766
|||||
Db 3071 ATTTTGTTAATGCAACAGGACTTAAATGAACCTTTGATCTGTGTTTTAAAGATTATTAA 3130
QY 2767 aaacattgttatctatatacatatgctcttgagagacttagcttcaactacactacagga 2826
|||||
Db 3131 AAAACATTTGCTATCTATACATATGGCTTGGAGACTTAGCTTCACTACACTACAGGA 3190
QY 2827 tatgatctccatgtagtccatataaaacctgcagagtgatttccagagtgctcgatctg 2886
|||||
Db 3191 TATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGTATTCACAGAGTGCCTCGATACTG 3250
QY 2887 ttaattacatctccattagggctgaaaagaatgacctacgttctctgtatatcacagtgtgt 2946
|||||
Db 3251 TTAATTTACATCTCCATTAGGGCTGAAGAATGACCTACGTTCTGTATACAGCTGTGTT 3310
QY 2947 gcttctgattgtgttactgttacacagaagtgtgtgcactgaagctctgcgtgtgtcc 3006
|||||
Db 3311 GCTTTTGATGTTGTGTACTGTACACAGAAAGTGTGTGCACGTGAGGCTCTGCGTGTGTC 3370
QY 3007 gtatgnaaaacctgtagccctgcaggttaagtaactgtctccattcatctgtttaacgtgg 3066
|||||
Db 3371 GTATGAAAACCTGTGAGCCCTGCGAGTTAAGTACTGCTTCATTCAATTGTTACGCTGG 3430
QY 3067 aattttctcccatggaatgtgaagtaaaacttaagtgttgcatacaataaatgtgaat 3126
|||||
Db 3431 AATTTTTCCTCCCATGGAATGTAAAGTTAAGTGTGTTCATCAATAAATGTTAAT 3490
QY 3127 actaaaaaaaaaaaaa 3142
|||||
Db 3491 ACTAAAAAAAAAAAAA 3506

RESULT 6
HSA132258 HSA132258 3066 bp mRNA PRI 14-MAY-1999
LOCUS
DEFINITION Homo sapiens mRNA for stauflen protein, partial.
ACCESSION AJ132258
VERSION AJ132258.1 GI:4572587
KEYWORDS stauflen gene; stauflen protein.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 3066)
TITLE Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
Interaction of influenza virus NS1 protein and the human homologue
of Stauflen in vivo and in vitro
JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE 99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS Ortin,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,
Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN
FEATURES
source
1. 3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
36. 1526
/gene="stauflen"
36. 1526
/gene="stauflen"
/function="dsRNA binding protein"
/product="stauflen protein"
/protein_id="CAB4082.1"
/db_xref="GI:4572588"
/translation="MKLGKKPMYKPPVDYSRMGSYNNYNNMRGAYPPRYFYPVPVPL
LYVELSVGGQFNGKGTROAAKHDAAKALRIQNEPLPERLEVNGRESEENLNK
SEISOVEFIALKRNLPNFNEVARRESGPHMKNFYKVSVEFVGEGRSKISKNA
AIVLEELKLPPLPAVERVYKPRIKKTKPIVKPQTSPEYGGINPISRLAQIQAKK
EKPEEYTLITERGLPRRREFYMQVGNHTAEGTGNKKVAKRNAENMELGFKVP
QRSHQRTQVRGEDTHKETRGWKKSNLFLNLALMENGTSNKEDEFMPYLSHQQLPA
GILPMVPEVAQAVGSQGHHTKDFTRAPNPAKATVTAMIARELLYGGTSPTAETLK
NNISSGHVPHGPLTRPSEQDLYLSRVQGFQVEYKDFPKNNKNEFVSLINCSQPLIS
HGICKDVESCHDMAALNLIKLSLSDOOSTEMPRRTGNPMSVCGRC"

BASE COUNT 891 a 693 c 699 g 783 t
ORIGIN

Query Match. 89.1%; Score 2799.4; DB 68; Length 3066;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2870; Conservative 0; Mismatches 6; Indels 7; Gaps 6;

QY 253 gcataccccctactgtagaactaaatgctgactgtgcatgaaacttggaaaaaaccaatgt 312
|||||
Db 1 GCATAACCCCTACTGTAGAACTAAATGCACCTGTGCATGAACCTTGGAACCAACCAATGT 60
QY 313 ataagcctgttaacccttactctcgtgagtcagttccacctataactacaacatgagagag 372
|||||
Db 61 ATAAAGCCTGTGACCTTACTCTCGGATGCAAGTCCACCTATAACTACAACATGAGAGAGAG 120
QY 373 gtgcttatcccccgaggtacttttaccattccagttccaccttacttatacaagtgg 432
|||||
Db 121 GTGCTTATCCCCCGAGGTACTTTTACCCTATTCACAGTTCACACTTACTTTATCAAGTGG 180
QY 433 aacttctgtgtagagagacgaactttaatgccaaggaagaaagacaagacaggtctggaaac 492
|||||
Db 181 AACTTTCTGTGGAGAGACAGCAATTTAATGGCAAAAGGAAGACAGACAGCTGCGAAGC 240
QY 493 acgatgtctgtgccaaagcgttgagatctcctgcagaatgagccctctgcagagaggtgg 552
|||||
Db 241 ACGATGCTGCTGCCAAAGCGTTGAGGATCTCCAGAAATGAGCCCTGCGACAGAGAGCTGG 300
QY 553 aggtgaatggaagaaatccgaagaagaaaaatcctaataatctgaaataagtcaagtgt 612
|||||
Db 301 AGGTGAATGGAAGAAATCCGAAGAAGAAATCTCAATTAATCTGAATAATTAAGTCAAGTGT 360
QY 613 ttgagattgcactaaacggaacttgcctgtgaaatttcgaagtggcccgagagagtggtcc 672
|||||
Db 361 TTGAGATTGCACCTTAAACGGAACCTTGCTGTGAATTTTCGAGGTGCGCCCGGAGAGAGTGGCC 420

KSVGEFVGESEKSKKISKKNAAIAVLELKLPLPAVERVKPRIKKTPIVRPO
TSPEYGGINPISRLAQIOQAKKEPEYTLTERGLPRREEVMQVGNHTAEGTG
TNKKVAKRNAENMLEILGFKVPQOPKPKSEKTPIKKPGDRKVTFEFGSD
ENGTSNKEDEFMRPYLSHQOLPAGILPMVPEVAQAVGSQGHHTKFTRAAPNPAKAT
VTAMIARELIXGTSPTAETILKNISSIGHVPHGLTRPSQDLYLSRVQFOVEYKD
FPKNKNFEVSLINCSSQPLISHGIGKDVESCHDMAALNIIKLSELDOOSTEMPT
GNPMSVCGRC"

BASE COUNT 965 a 875 c 790 g 876 t
ORIGIN

Query Match 92.2%; Score 2896; DB 10; Length 3506;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 cagaaagcataaccctactgtagaactaaatgcactgtgcatgaaacttggaaaaaac 306
Db 611 CAGAAAGCATAACCCCTACTGTAGAATAATGCACCTGTGCATGAAACTTGGAAAAAAC 670
QY 307 caatgtataagcctgttgacccttactctcggatgcagtcacactataactcaacatga 366
Db 671 CAATGTATAAGCCTGTGTACCTTACTCTCGGATGCAGTCCACCTATAACTTACAACATGA 730
QY 367 gaggaggtgctatcccccgagtaacttaccatccagtlccaccttacttatac 426
Db 731 GAGGAGGTGCTTATCCCGCAGTACTTTTACCCTATTCAGTTCACCTTACTTTATC 790
QY 427 aagtgaaacttctgtgagagacagcaatttaatgycaaaagaaagacaagacagctg 486
Db 791 AAGTGGAACCTTCTGTGGAGGACAGCAATTTAATGCCAAAGGAAAGACAAGACAGGCTG 850
QY 487 cgaacaacgatgtctgtgccaaagcgtltgagatcctgcagaaatgagccctgcagaga 546
Db 851 CGAAACGCGATGCTGTGCCAAAGCGTGTAGGATCCTGCAAGATGAGCCCCCTGCCAGAGA 910
QY 547 ggcctgaggttgaatggaagaaatccgaagaagaaatctcaataatctgaaataagtc 606
Db 911 GGCTGGAGGTGAATGGAAGAGAAATCCGAAGAAGAAATCTCAATAATCTGAAATAAGTC 970
QY 607 aagtgcttgaagtgtcacttaaacggaacttgcctgtgaaattcgaggtgcccggaga 666
Db 971 AAGTGTGAGATTGCACTTAAACGGAACTTGCTGTGAATTTGAGGTGGCCGGGAGA 1030
QY 667 gtgcccaccccacatgaagaacttltgaccaagtttcggttgggagtttgtgggg 726
Db 1031 GTGCCCAACCCACATGAAGAACTTTGTGACCAGGTTTCGGTTGGGAGTTGTGGGGG 1090
QY 727 aagtgtaagggaaaaagcaagaagatttcaaaagaaatgcccacatagctgttcttgag 786
Db 1091 AAGGTGAAGGGAAGCAAGAGATTTCAAGAAGAAATGCCGCATAGCTGTCTTGAGG 1150
QY 787 agctgaagaagtttaccgcccctgcctgcagtgtgaacgagtaaaagcctagaatacaaaaga 846
Db 1151 AGCTGAAGAAGTTACCGCCCTGCTGCTGAGTTGAACGAGTAAGCCTTAGAATCAAAAAAGA 1210
QY 847 aaacaaaaccatagtcgaagccacagacaagcccagatatgycaggggatcaatccga 906
Db 1211 AAACAAAACCCATAGTCAAGCCACAGAACGCCAGAAATATGGCCAGGGATCAATCCGA 1270
QY 907 ttagcgcactgccccagatccagcagycaaaagagaaagagagccagagtacacgctcc 966
Db 1271 TTAGCCGACTGGCCCGACAGATCCAGCAAGCAAAAAAGAGAGAGCCAGAGTACACGCTCC 1330
QY 967 tcacagagcagagcctcccgccgcccagggaggttctgtatgcatgaggttggaaacc 1026
Db 1331 TCACAGAGCGAGGCCCTCCCGCCGCCAGGAGTTTGTATGCAGGTGAAGGTTGGAAGACC 1390
QY 1027 acactgcaagaaggaacgggcaaccaacaagaaggttgcccaagcgcgaatgcaagcgaaga 1086
Db 1391 ACACCTGCAGAGGAAGGACGGCACCAAGAAGGTGGCCCAAGCGCAATGCAGCCGAGAGACA 1450
QY 1087 tgcctgagatccttggtttcaaaagtcctcgagcgcgagcccaaccaaacccgcactcaagt 1146
|||||

Db 1451 TGCTGAGATCCTTGGTTTCAAGTCCCGCAGCGGCAGCCACCAAAACCCGCACTCAAGT 1510
QY 1147 cagagagagaagacaccccataaagaaaaccagggtatggaagaaagtaaccttltgaa 1206
Db 1511 CAGAGGAGAAGACACCCCATAGAAGAACCGAGGGATGGAAGAAAGTAACCTTTTGAAC 1570
QY 1207 ctggtcttgggatatgaaaatgaggactagtaataaagagatgagttcagatgacctatc 1266
Db 1571 CTGGCTCTGGGGATGAAAAATGGGACTAGTAATAAGAGGATGAGTTCAGAGATGCCTTATC 1630
QY 1267 taagtaatcagcagctgcctgcctggaattcttcccatgltgccgaggtgcgccagctg 1326
Db 1631 TAAGTCAATCAGCAGCTGCTGCTGGAATTTCTCCATGGTGCCTGAGGTGCCCCAGGCTG 1690
QY 1327 tagagttagltcaagagacatcacaccaagaalltlaccaggycagctccgaatccgtcca 1386
Db 1691 TAGAGTTAGTCAAGGACATCACACCAAGATTTTACAGGGCAGCTCCGAATCCTGCCA 1750
QY 1387 aggcacggtlaactgcatgataagcccgagagttgtgtatggyggcacctcgccacag 1446
Db 1751 AGGCCAGGTAACCTGCATGATAGCCCCGAGAGTGTGTATGGGGCACCCTGCCACAG 1810
QY 1447 ccgagaccattttaaagaatacatctcttcaggccacgtlaccocatgacctctca 1506
Db 1811 CCGAGACATTTTAAAGATATACATCTCTCAGGCCACGTACCCCATGACCTCTCACGA 1870
QY 1507 gacctctgaagcaactggaactatcttccagagtlccagggttccaggttgaatacaaa 1566
Db 1871 GACCCCTGAGCAACTGAGACTTACTTTCCAGAGTCCAGGGATTCCAGGTTGATACAAAG 1930
QY 1567 acttccccaaaaacaacaagaacgaaattgtatctcttatacaattgtcctctcagccac 1626
Db 1931 ACTTCCCAAAAAACAACAAGAACTTGTATCTCTTATCAATTTGCTCCTCAGCCAC 1990
QY 1627 ctctgatacgccatggtatcggcaagatgtgagttcctgcatagatatgctgcgtga 1686
Db 1991 CTCTGATCAGCCATGTTATCGGCAAGGATGTGAGTCTGCCATGATATGCTGCGCTGA 2050
QY 1687 acatcttaaaagttgtctgtctgagltggaaccaacaagtacagagatgccaagaacaggaa 1746
Db 2051 ACATCTTAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGACAGGAA 2110
QY 1747 acggaaccaatgtctgtgtgtggaaggtgtgtgaaccttltctgccaatgaaccattataa 1806
Db 2111 ACGGACCAATGTCTGTGTGTGGAGGTGCTGAACCTTTTCTGCCCATGAACCATTAATAA 2170
QY 1807 atcccaacatatatactgaanaatactgaaactgcttgaanaaatttgaatttctgatacc 1866
Db 2171 ATCCCAACATATATACTGAAGAAATACTGAAACTGCTTTGAAAAATTTGGAATTTGATACC 2230
QY 1867 tccagtgyccgagagacacggttggtaaagatgtgycagcagcaggygaagacaacag 1926
Db 2231 TCCAGTGGGCCGAGAGACACGGTGGGTAAGGATGTGGGCAGACAGGAGAACACAACAG 2290
QY 1927 aaacacaaggaagcgcgtgtgtgcccgcgtgactgtgtctgggttltgtgtatgycact 1986
Db 2291 AAACACAAGAGAGGGCGGTGTGTGCCCGGCTGACTGTGTGTGGGTTTGTGTGATGGCCACT 2350
QY 1987 cgttgacctgycgtlccctacgcaatagcagctgcctgttggggaagaagagctgcccagc 2046
Db 2351 CGGTGACTGGCGGTCCCTACGCAATAGCAGCTGCTGTGGGGAAGAGGGCTGCCACAGC 2410
QY 2047 cagctgttctcccgagacacagcagatccacaccccttggcacctccgttltgtlct 2106
Db 2411 CAGCTGTTCTCCCGGAGACACCGAGATCCACACCCCTGGGCACCTCCGTGTGTGCTT 2470
QY 2107 ttttccctgtgtgaagaagaagaacgycagaccccttcaagctgtgtcaactcaga 2166
Db 2471 TTTTTCCTCTGTGTGAAGAAGAAACGCGACGACCCCTTCTCAAGCTGGCTCACTCAGA 2530
QY 2167 cacatlggacaacacccctlgacagccatgcccagagagagcccttgaacgycgccagagc 2226
Db 2531 CACATTTGGACAACCCCTGGACAGCCATGCCAGAGAGAGGCCCTTTGACCGGCCACAGAGC 2590

Db 1801 TTGTATCTCTTAJCAATTGCTCCTCTCAGCCACCTCTGATCAGCCANIGTATCGGCAAGG 1860
QY 1654 atgtgagtccttgccatgatatgctgcgtgaaacatcttaagtgtgtctgagttgg 1713
Db 1861 ATGTGGAGTCCCTGCCATGATATGGCTGCGCTGAACATCTTAAGTGTGCTGTGAGTTGG 1920
QY 1714 accaacaagtlacagagatgccaagaacaggaacacggaaccaatgtctgtgtgtggaggt 1773
Db 1921 ACCAACAAAGTACAGAGATGCCAAGAACAGGAACGACCAATGTCTGTGTGGAGGT 1980
QY 1774 gctgaaccttcttgccatgaaacatataaaalcccaacatatatactgaaatactg 1833
Db 1981 GCTGAACCTTTCTGGCCATGAACCATTTATAAATCCCAACATATATTAAGTGAATACTG 2040
QY 1834 aaactgcttgaanaatttgnaatttctgtatacctccagtgaggccgagagacaagtgaggt 1893
Db 2041 AAACCTGCTTTGAAAAATTGGAATTTCTGATACCTCCAGTGGGCCGAGACACGGTGGGT 2100
QY 1894 aaagatgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1953
Db 2101 AAAGGATGTGGGCAGCAGCAGGGAAGACACAGAAACACAAAGAGAGCGCGCTGTGGCCGGC 2160
QY 1954 tggactgtgtgaggt 2013
Db 2161 TGGACTGTGTGGGGT 2220
QY 2014 gcagctgcctgtgagggaagaagagcgtgcccagcagcagcgtgtctcccgaggacacagcag 2073
Db 2221 GCAGCTGCCGTGGGGAAGAAGGGCTGCCGCCACGCTGCTCTCCCGGACACAGCAGCAG 2280
QY 2074 atccacacccctgggcacccctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2133
Db 2281 ATCCACACCCCTGGGCACTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
QY 2134 ggcacgaccccttctcaagctgtgtcactcaacacacacacacacacacacacacacacacac 2193
Db 2341 GGCACGACCCCTTCTCAAGCTGGCTCAGCTCAGACACACATTTGGGACAAACCCCTGGACAGCCA 2400
QY 2194 tgcagagagagccttctgaccggcccccagagcctaagaacacacacacacacacacacacac 2253
Db 2401 TGCCAGAGAGAGGCGCTTTTGACCGCGCCCGCAGACTAAAGACACAGAGAAATCAAAATGCT 2460
QY 2254 tccctactcagcgtgacccaacttctctagtgccacgccccacacacacacacacacacacac 2313
Db 2461 TCCTACTCAGCGTGACCCAACTTTCTAGTGCCACGCGCCACACCACTCCCTGCAGTAC 2520
QY 2314 ccacacatcacccac 2373
Db 2521 CCACACCATCACACACCTTCTCTCTCCACACAGTGATCTGTATCTTAGTTTCATTATTT 2580
QY 2374 tctttgatgtatgac 2433
Db 2581 TCTTTGATGTATGACACTATATAAATTTTCATTGAGAAATTTCTCAATTTGTATCTTA 2640
QY 2434 gttaaatagcacagcttggaacactgtctcagacactgacttatacaataactcaaccgaca 2493
Db 2641 GTTAATAGCACAGTTTGGAAACTGTCTGAGACTGACTTTATCAATTAATCTAACCGACA 2700
QY 2494 aagatcatatccatgtatgtgtgttagacacacacacacacacacacacacacacacacacacac 2553
Db 2701 AAGATCATATCCATGTCTATGTGTAGACATTTTATTTTCATTGACTAACCCAGGACAG 2760
QY 2554 ttccagtgatgcaaatgt 2613
Db 2761 TTTCAGTGATGCAAAATGTGTGCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2820
QY 2614 accctgaataagctcccccacagctgtatataaatttgacaacacacacacacacacacacacac 2673
Db 2821 ACCTTGAATAGCTCTCCACACAGTGTATATAATTGACACATTTAGCAATTTTAAACTTTAG 2880
QY 2674 atgatcatctgtgtccac 2733
Db 2881 ATGATCATTTGCTTCATTTTATTTATTTTATTTTATTTTGTAAATGCAACAGGACTTAA 2940

QY 2734 atgaacattgatctctgtgtttaaagattatataaaaaacacacacacacacacacacacacac 2793
Db 2941 ATGAACTTTGATCTCTGTTTAAAGATTATTAATAAACATTTGTGTATCTTATACATATAGGC 3000
QY 2794 tcttgaggacttagcttctcaactac 2853
Db 3001 TCTTGAGGAGCTTAGCTTTCACCTACACTACAGAGATGTGATCTCCATGTAGTCCATATAAAC 3060
QY 2854 ctgcagagtgatcttccagagtgctcgaatactgttaattacatctccataggcgtgaaa 2913
Db 3061 CTGCAGAGTGATTTTCCAGAGTGCCTCGATACGTGTTAATTACATCTCCATTAGGGCTGAAA 3120
QY 2914 agaatacctacgcttctcgtatacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2973
Db 3121 AGAATGACCTACGTTTCTGTATACAGCTGTGTGCTTTTGATGTGTGTGTGTGTGTGTGTGT 3180
QY 2974 gaagtgtgtcactgagcctcgt 3033
Db 3181 GAAGTGTGTGCACGTAGGCTCTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3240
QY 3034 ttaagtactgcttccatcatctgtttaaagctggaatttctcccatggaatgaagta 3093
Db 3241 TTAAGTACTGCTTCATTCATTTGTTACGCTGGAATTTTCTCCCATGGAATGTAAGTA 3300
QY 3094 aaacttaagtgttgtlcatcaataaatgttaataactaaataaaaaaa 3142
Db 3301 AAACCTTAAGTGTGTGTATCAATAATGTAATACTAAAAA 3349
RESULT 5
AF061939
LOCUS AF061939 3506 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced, complete cds.
ACCESSION AF061939
VERSION AF061939.1 GI:4335948
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3506)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE Mammalian stauflen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3506)
AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1. .3506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. .3506
/gene="STAU"
409. .2142
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="stauflen protein"
/protein_id="AADI7531.1"
/db_xref="GI:4335949"
/translation="MSQVOYOVONPSAALSGQILNKNQSLSQPLMSIPTSTSLPS
ENAGRIQNSALPSASITSTSAAESITPTVELNALCMKLGKPMYKPVDPYSRMOST
VYNMRGGAAPRYEYFPVPVPLLYQVELSVGGQFNGKGTROAAKHDAAKALRLI
QNEPLPERLEVNGRESEENLNKSEISOVEIALKRNLPVNFVARESGLPHMKNFVT

QY 1516 agcaactgactatcttccagagtcaggagaltccaggttgaatatacaagacttccca 1575
Db 1558 AGCAACTGGACTATCTTCCAGAGTCCAGGGATTCCAGGTGGAATACAAAGACCTCCCA 1617
QY 1576 aaacaacaagaacgaatttgtatctctatacaattgtctcctcagccactctgatca 1635
Db 1618 AAACAACAAGAAGCAATTGTATCTCTTATCAATGTCTCTCCAGCCACCTCTGATCA 1677
QY 1636 gccatgtatcggcaagatgtgtgagtcctcgcatgatatgtgcgtgaacatcttaa 1695
Db 1678 GCCATGGTATCGGCAAGATGTGAGTCCCTGCCATGATATGGCTGCGGTGAACATCTTAA 1737
QY 1696 agttgctgtctgagttggaaccaaaagtatacagagatgccaagaacagaaacgaa 1755
Db 1738 AGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAAGAAACGGACCAA 1797
QY 1756 tgtctgtgtgtgagggtgctgaaccttctcggccatgaaccaltataaatcccaaca 1815
Db 1798 TGTCTGTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAACCATTTATAAATCCCAACA 1857
QY 1816 tatactgaaaaaactgaacctgttgaaaatttgaatttctgatacctccagtgagg 1875
Db 1858 TATATACTGAAAAATACTGAAACTGCTTTGAAAAATTGGAATTCTGATACCTCCAGTGGG 1917
QY 1876 ccgagagacacggtgtggttaaagatgtgtggcagcagcagggaaagacacagaacaaga 1935
Db 1918 CCGAGAGACACGGTGGGTAAAGGATGTGGGCAGCAGCAGGGAAGACACAGAAACACAAG 1977
QY 1936 gaagcgctgttgcc-ggctgtgactgtgtgggttctgtgtgatgtgccaactcggtgacc 1994
Db 1978 GAGCGGCTGTGGCCGGGCTGGACTGTGCTGGGTTTGTGTGATGGCCACTCGGTGACC 2037
QY 1995 tggcggtccctaagcaatagcagctgcctgtgtgggaagaaggcctgcccagcagctggt 2054
Db 2038 TGGCGGTCCCTACGCAATAGCAGCTGCCTGTGGGAAGAGGGCTGCCAGACGACTGGT 2097
QY 2055 tctcccggaacaccagcagatccacaccctggcgacactccgtgtgtgtcttttttcc 2114
Db 2098 TCTCCCGGACACCCAGCAGATCCACCCCTGGGACCTCCGTGTTGTGCTTTTTC 2157
QY 2115 cctgtgtgaaagaagaacggaacgcaaccccttctcaagctgtgtcactcagacacattg 2174
Db 2158 CCTGTGTAAAGAAGAACGGCAGACCCCTTCTCAAGCTGGCTCAGACACATTTGG 2217
QY 2175 gacaacccctgacagccatgcagagagagagggccttgaaccggccccagagctaaagca 2234
Db 2218 GACAAACCTTGACAGCCATGCCAGAGAGAGGCTTTGACCGGGCCCGAGAGCTTAAAGCA 2277
QY 2235 ccagagaaaaataaatgtctcactcagcgltgaccaacttctagtgtccaagcgcc 2294
Db 2278 CCAGAGAAATCAAAATGCTTCCTACTCAGCGTGACCCCACTTTCTAGTGTGCCACGGCC 2337
QY 2295 ccaccacctcctgagtagcaccaacatcaccaactgtctctctccaacagtgatcgt 2354
Db 2338 CCACCACTCTCTGACGTACCCACACCATCACCACTGCTTCTCTTCCAACAGTACTGT 2397
QY 2355 attcttagttcattatttcttttgattgatatagacactatataaaatttcaatttgag 2414
Db 2398 ATTCTTAGTTTCATTAATTTCTTTTGATGTGATATGACACTATATAAATTTTCATTGGAG 2457
QY 2415 aatttcacaattgtatctagttaataatagcacaglttggaaaactgtctgagagtgactt 2474
Db 2458 AATTTCATAATTGTATCTAGTTAATAATAGCACAGTTTGGAAACTTGTCTGAGACTGACTTT 2517
QY 2475 atcaataatctaaccgacaagaatataccatgtlatagtlgttagacatttatttc 2534
Db 2518 ATCAATAATCTAACCAGACAAGATCATATCCATGTGTATGTGGTTAGACATTTTATTTC 2577
QY 2535 attgactaaccagagacaglttcagltgatgcaaatgtgtccctcgtgtcagctga 2594
Db 2578 ATTGACTAACCCAGGACAGTTCAGTGATGCAAAATGTGTGCCCTCTGCTCAGCTGA 2637

QY 2595 cagtccctgacttcaaaaaccttgaataagtcctcccaagltgtataaatltgacaatt 2654
Db 2638 CAGTCCCTGGACTTTCAAAAACCTTGAAITAAGTCTCCACAGCTGTATATAATTGGACAATT 2697
QY 2655 taggaattttaaactttagatgatcatlttggttccatttatttcaatttatttltgt 2714
Db 2698 TAGGAATTTTAAACTTTAGATGATCATTTGGTTCATTTTATTTCATTTTACTTTTGT 2757
QY 2715 taatgcaacagagactlaaatgaacttgcctcgttlttaagaaltataaaaaacalt 2774
Db 2758 TAATGCAACAGGACCTTAAATGAACCTTTGATCTCTGTGTTTAAAGATTATTAAAAACATT 2817
QY 2775 gtgtatctatacatatggtccttgaaggacttagcttccaactacactacagatatgact 2834
Db 2818 GTGTATCTATACATATATGGCTCTTGAGGACTTAGCTTTCACACTACAGAGATATGATCT 2877
QY 2835 ccatgtatccatataaaacctgcaagtgatlttccagagtgctcgatactgttaattac 2894
Db 2878 CCATGTAGTCCATATAAACCCTGCAGAGTGATTTTCCAGAGTGCTCGATACTGTTAATTAC 2937
QY 2895 atctccaltagggtcgtgaaaagaatgacctaagtttctgtatatacagctgtgtgtcttga 2954
Db 2938 ATCTCCATTAGGGCTGAAAAGAAATGACCTACGTTTCGTATATACAGCTGTGTGCTTTTGA 2997
QY 2955 tgttgttactgttacacagaagtgtlgtcactgaagcctcgtcgtgtgtccgtatggaa 3014
Db 2998 TGTGTGTACTGTACACAGAGTGTGTGCACTGAGGCTGTGCGTGTGTCCGTATGGAA 3057
QY 3015 aacctgtgagccctgagagtaagtaactgtctccatcatcattgttlaagctgtgaatttcc 3074
Db 3058 AGCCTGTAGCCCTCGAGAGTAACTACTGCTTCATTCATTGTTTACGCTGGAATTTTTC 3117
QY 3075 tccccatgyaatgtlaagtaaaacttaagtlttgtcatcaataaatgtgtaataactaaaa 3134
Db 3118 TCCCCATGGAATGTAGTAAGACTTAAGTGTGTTGTCAATCAATAATGTGTAATCTAAAAA 3177
QY 3135 aaaaaaaa 3142
Db 3178 AAAAAAAA 3185

RESULT 4
AF061941
LOCUS AF061941 3349 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced, complete cds.
ACCESSION AF061941
VERSION AF061941.1 GI:4335952
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3349)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.
TITLE Mammalian stauflen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3349)
AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source location/Qualifiers
1..3349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1..3349
/gene="STAU"
495..1985
CDS

REFERENCE	1 (sites)
AUTHORS	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE	NEDO human cDNA sequencing project
JOURNAL REFERENCE	Unpublished (2000)
AUTHORS	2 (bases 1 to 3194)
TITLE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
JOURNAL	Direct Submission
COMMENT	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
FEATURES	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
source	Location/Qualifiers 1..3194 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="HepG2" /cell_type="hepatoma" /clone="HEP02379" /clone_1lb="HEP" /note="cloning vector pME18SFL3"
misc_feature	1..3194 /note="highly similar to AF061938 Homo sapiens stauferin protein (STAU) mRNA"
BASE COUNT	894 a 787 c 732 g 781 t
ORIGIN	
Query Match	93.7%; Score 2945.2; DB 34; Length 3194;
Best Local Similarity	97.1%; Pred. No. 0;
Matches 3094; Conservative	0; Mismatches 13; Indels 81; Gaps 6;
Dn	33 ctacagcgtttgcgc--ggcgctgcgctctctctcgcgtccgcgttcttccttgaccgcc 90 1 CTTCAGCGTTGGCGCGGCGCTGCCGCGTCTCTCGGCTCCGCTTGTGACCGCC 60
Oy	91 tcccccccccggccggcgcgccgcgcctctctcaacgagccaatcgcctcttcctccct 150 61 T-CCCCCCGCGCGCGCGCGCCTCCTCCACGCGCACCTCCGCTTCCCTCCCT 119
Dn	151 tcgtccctctctcctcctcccttttctctctctctccttcctccctcgcgcgaacgcc 210 120 TCCTCCCTTCTCTCTCTCCCTTTTTCCTTCTCTCTCCCTCCGCGCACCGCCA 179
Oy	211 ggaccggcgccgggggagcagctcgagcagcagcc----- 247 180 GGACCGCGCGCGCGGAGCACTCGAGCAGCAGCCAGAGTTATTAACTTAACCTTAACCT 239
Oy	248 -----agaagca 255
Dn	240 CTCAGAAGTGAACAAGACACATTTGTTCTTGGAACGCCCTTTTATAAAAAAAGAACGA 299
Oy	256 taaccacctactgtagaactaaatgcactgtgtcatgaaacttggaaaaaaaaccaatgtata 315 300 TAACCCCTACTGTAGAACTAAATGCACCTGTGCATGAAACTTGAAAAAACCAATGTATA 359
Oy	316 agcctgttacacctactctcgatgcagtcacacctataactacaacatgagagagtgt 375 360 AGCCTGTGACCCTTACTCTCGCATGCAGTCACCTATAACTAACATGAGAGAGAGGTG 419
Oy	376 ctatcccccgaggtactttaaccattccagttccaccttacttatcaagtgaac 435

Db	420	CTTATCCCCCGAGGTACTTTTACCATTTCACGTTCCAGTCCACCTTTTACTTTATCAAGTGAAC	479
QY	436	tctctgtgggagagcagcaatttaatgycaaagaaagacaagcagctgcgaaacg	495
Db	480	TTTTCTGTGGAGAGACAGCAATTTAATGGCAAAGGAAGACAAGACAGCTGGCAACACG	539
QY	496	atgctgtctgcctcaagcgtttgagatccttcgcaaatgagccccctgcagagagcgtcygag	555
Db	540	ATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGATGAGCCCCCTGCCAGAGAGGCTGGAGG	599
QY	556	tgaatyggaagagaatccggaagaagaaaaatctcaataaactgaaataagtcgaagtgttg	615
Db	600	TGAATGGAAGAGAAATCCGAGAAGAAATCTCAATAAATCTGAATAAAGTCAAGTGTG	659
QY	616	agattgcactlaaacggaacttgctgtgaatttcgaagtgcgcggagagatgcccac	675
Db	660	AGATTGCACCTTAAACGGAACTTGCTGTGAATTTGAGGTGGCCCGGAGAGTGGCCAC	719
QY	676	cccatatgaagaactttgtgaccaaagtttcggttggtggagttgttggggaaggtgaag	735
Db	720	CCCACATGAAGAACCTTGTGTGACCAGGTTTCGGTTGGGAGTTTGTGGGGGAAGGTGAAG	779
QY	736	ggaaaagcaagaagatttcaaagaaaaatgcgcgcatagtgttcttgagagctgaaga	795
Db	780	GGAAAGACGAAGAAGATTTCAAAGAAATGCCCCCATGCTGTCTTGAGGAGCTGAGAA	839
QY	796	agttaccgccccctgtcgtgcagttgaaagagtaagccttagaatcaaaaaagaaacaaac	855
Db	840	AGTTACCGCCCCCTGCTGCAGTTGAACGAGTAAAGCCTTAGAATCAAAAAAGAAACAAAA -	898
QY	856	ccatagtcaagccacagacaaagcccagaatatatgcccagggatcaatccgatlagccgac	915
Db	899	CCATAGTCAAGCCACAGACAGACCAGAAATATGCCCCAGGGGATCAATCCGATTAGCCGAC	958
QY	916	tggccccagatccagcagcgaagaaagagaaagagcgaagtaacgctcctcacagagc	975
Db	959	TGGCCCAAGATCCAGCAGCGCAAAAAAGAGAAGAGCCCAAGTACACGCTCCTCACAGAGC	1018
QY	976	gaggcctcccgcgccgcaaggagtttgtatgcaggtgaaagtttgaaaccacactgcag	1035
Db	1019	GAGGCTCCCCCGCGCGCAGGAGTGTGTGATGCAGGTGAAGGTTGGAAACCACACTGCAG	1078
QY	1036	aaggaacggygcaccacaagaaggttgcaccaagcgcaatgycagccgagaacatgtctgaga	1095
Db	1079	AAGGAACGGGCACCACAGAAGGTGGCCAAAGCCCAATGCAAGCCGAGAAGACATGCTGAGA	1138
QY	1096	tccttgtttcctaaagtcctccgacgycgagcccaaccacccgcactcaagtccagagaga	1155
Db	1139	TCCTTGGTTTCAAAAGTCCCGCAGGCGCAGCCCAACCAACCCGCACCTCAAGTCAGAGAGA	1198
QY	1156	agacacccataaagaacacagggatggaagaaagtaaccttlttgaacctgctctg	1215
Db	1199	AGACACCCCATAAAGAAACCA - GGGATGGAAGAAAAAGTAACTTTTGTGAACCTGGCTCTG	1257
QY	1216	gggatgaaaatggyactagtaataaagagatgagttcagagatgcttatctaagtcac	1275
Db	1258	GGGATGAAAAATGGGACTAGTAAATAAGAGGATGAGTTCAAGATGCCCTTATCTAAGTCATC	1317
QY	1276	agcagctgcctgtgtaatttcccatgtgtgccgaggtgcgccaggtgtagagta	1335
Db	1318	AGCAGCTGCCTGCTGGAATTTCTCCCATGTGTGCCGAGGTGCCCCAAGCTGTAGAGTTA	1377
QY	1336	gtcaagyacatcacaccaaagatttaccagggcagctccgaatcctctgcacaagccaag	1395
Db	1378	GTCACAGGACATCACACCACCAAGATTTTACCAAGGGCAGCTCCGAATCCTGCCAAGGCCACGG	1437
QY	1396	taactgccatgatagcccgagagttgttgtatggyggcactcgcccaagccgagacca	1455
Db	1438	TAACTGCCATGATAGCCCCGAGAGTTGTTGTATGGGGGCACCTGCCCCACACAGCCGAGACCA	1497
QY	1456	ttttaaagaataacatctcttcagggccacgtaccaccatygaaactctcagagaccctctg	1515
Db	1498	TTTTTAAAGATTAACACCTCTTCAGGGCCACGTACCCTCATGCACTCTCACAGAGACCTCTG	1557

QY 1186 gaagaagtaaccttttttgaaactggtctctgggatgaaaaatgagactagtaataaagag 1245
|||||
Db 1261 GAAAGTAACCTTTTGTGAACCTGGCTCTGGGATGAAAAATGGACTAGTAATAAAGAGG 1320
QY 1246 atgagttcaagatgacctatcaagtcacagcagctgctgtgaattcttcccatgg 1305
|||||
Db 1321 ATGAGTTCAAGGATGCTTATCTAAGTCATCAGCAGCTGCTGCGAATTCTCCCATGG 1380
QY 1306 tggccgaggtcgcccaagctgtagagttagtcaagagacatcacaccaagatttaacca 1365
|||||
Db 1381 TGCCCGAGGTGCCCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGATTTTACCA 1440
QY 1366 gggcagctccgaatccctgcgaagccaagtgtaactgcatgtagcccgagagtgtgt 1425
|||||
Db 1441 GGGCAGCTCCGAATCTGCGCAAGGCCACGGTAAC TGCCATGATAGCCCGAGAGTGTGT 1500
QY 1426 atggggcaccctcgcccaagccgagaccattttaagaataacatctcttcagggccag 1485
|||||
Db 1501 ATGGGGGACCTCGCCACAGCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCACG 1560
QY 1486 taccccaatgacctctcaacgaacccctctgagcaactgactatcttccagagtcagg 1545
|||||
Db 1561 TACCCCATGGACCTCTCACGAGACCTCTGAGCAACTGACTATCTTTCAGAGTCCAGG 1620
QY 1546 gatccaggttgaatacaaaagctcccaaaacaaacaagaagaaatltgtatctcta 1605
|||||
Db 1621 GATTCAGGTGAATACAAAGACTTCCCAAAAAACAACAAGCAATTTGATCTCTTA 1680
QY 1606 tcaatgtcctctcagccacctctgatacagcatgtgtatcgcaagatgtgagtcct 1665
|||||
Db 1681 TCAATGTCTCTCTCAGCCACCTTGATCAGCCATGGTATCGGCAAGATGTGAGTCTT 1740
QY 1666 gccatgatatgtctgcgtgaacatcttaagtctgtctgtgagttgaccaacaagtta 1725
|||||
Db 1741 GCCATGATATGGCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGGACCAACAAGTA 1800
QY 1726 cagagatgccaagaagaacaggaagaagccaatgtctgtgtggaggtgtgtaaccttt 1785
|||||
Db 1801 CAGAGATGCCAAGAAGACAGGAAGGAGCAACCAATGTGTGTGTGGAGGTGCTGAACCTTTT 1860
QY 1786 ctggccatgaaccatlaaatacccaacatatatactgaaaaatactgaaactgttga 1845
|||||
Db 1861 CTGGCCATGAACCATTTATAAATCCCAACATATATAGTGAATACTGAACCTGCTTGA 1920
QY 1846 aaatttgaattctgataccctcagltggcgagagacacggttgggtaaagatgtgg 1905
|||||
Db 1921 AAATTTGGAATTTCTGATACCTCCAGTGGCGGAGAGACACGGTGGGTAAAGGATGTGGG 1980
QY 1906 cagcagcaggggaagacaacagaaaacacaagagcggtgtgtgcccgtctgactgtgt 1965
|||||
Db 1981 CAGCAGCAGGGAAGACACACAGAACAACAAGAGCGGCTGTGGCCGGCTGAGCTGTGCTG 2040
QY 1966 gggtttgttgtatgycgcaactcgtgtgacctgycgtccctacgcaatagcagctgtct 2025
|||||
Db 2041 GGGTTGTTGTGATGGCCACTGCGGTGACCTGGCGGTCTTACGCAATAGCAGCTGCTGT 2100
QY 2026 ggggaagaagggtctgcccagccagctgttctcccggaacaaccagcagatccacacctg 2085
|||||
Db 2101 GGGGAAGAAGGGCTGCCACGACGCTGTTCTCCGGGACACCAAGCAGATCCACACCTTG 2160
QY 2086 ggcacctccgtgttgtgtcttttttccctgtgtgtaagaagaagaacgacgacacct 2145
|||||
Db 2161 GGCACCTCCGTGTGTGCTTTTTCCTCTGTGAAGAAGAACGGCAGCACCCT 2220
QY 2146 tctcaagctgtctactcagacacatttgggacaaaccttgacagccatgtccagagagag 2205
|||||
Db 2221 TCTCAAGCTGGCTCACTCAGACACATTGGGACAAACCTGGACAGCCATGCCAGAGAGAG 2280
QY 2206 gaccttgaccggccccagagctaaagaacaccagagaaaaatcaaatgttctctaactcagcg 2265
|||||
Db 2281 GCGTTTGACCGGGCCAGAGCTAAAAAGCACACAGAGAAAAATCAAAATGCTTCTACTCAGCG 2340
QY 2266 tgaccacaactttctagtgtgtccaacgccccaccacctcctgtagtaccacacatcac 2325

Db 2341 TGACCCAACCTTTCTAGTGTGCCAGGGCCCCACACACCTCTGCAGTACCCACACATCAC 2400
QY 2326 cactgcttctctcccaacagtgatctgtattcttagtttcaattatttcttcttgatga 2385
|||||
Db 2401 CACTGCTTCTCTTCCACACAGTGATCTGTATTTCTTAGTTCAATTATTTCTTTGATGTA 2460
QY 2386 tatgacactatataaaatttcaatttgagaatttctcaatttgtlatctagttaataagcac 2445
|||||
Db 2461 TATGACACTATATAAATTTTCATTTGAGAAATTCTCAATTGTATCTAGTTAAATAGCAC 2520
QY 2446 agtttgaaaactgtctgagactgaccttatacaataatctaaccgacaagaatcatatcc 2505
|||||
Db 2521 AGTTTGAACCTGTCTGAGACTGACTTTATCAATATCTAACCAGCAAGATCATATCC 2580
QY 2506 atgtgatgtgttagacatttatttcaatttgactgaactaaccagagacagtttcagtatgc 2565
Db 2581 ATGTGATGTGGTTAGACATTTTATTTTATTTGACTTAACCCAGACAGTTTCAGTATGC 2640
QY 2566 aaattgtgccccctgtgtcagctgaaacagtcctggaacttcaaaaaccttgaataag 2625
Db 2641 AAATGTGTGCCCTCTGTTCAAGCTGAACAGAGTCCGTGACTTCAAAAAACCTGATATAAG 2700
QY 2626 tctcccaagttgtataaatgtgacaattttaggaattttaaactttagatgatcatttg 2685
Db 2701 TCTCCACAGTTGTATAAATTGGACAATTTAGGAATTTTAACTTTAGATGATCATTTGG 2760
QY 2686 ttccatttatttcaatttatttlttltgtlaatgcaaacaggaacttaaatgaactttgat 2745
Db 2761 TTCCATTTTATTTTCAATTTTATTTTGTTAATGCAACAGGACTTAAATGAACCTTTGAT 2820
QY 2746 ctctgtttaaagaattatlaaaaaacattgtgtatctatataatgctcttgaagacct 2805
Db 2821 CTCTGTTTAAAGATTATTAATAAACATTGTGTATCTATACATATGCGCTTGAGGACTT 2880
QY 2806 agcttcaactacactacagagatagatctccatgtagtcataataaacctgcagagtgtat 2865
Db 2881 AGCTTTCACACTACACTACAGAGATGATCTCCATGTAGTCATATAAACCCTGCAGAGTGAT 2940
QY 2866 ttccagagtgctgatactgttaattacatctccattagggctgaaagaatgacctac 2925
Db 2941 TTTCCAGAGTGCCTGATACTGTATATTACATCTCCATTAGGGCTGAAAAAGATGACCTAC 3000
QY 2926 gttctgtatacagctgtgtgttctttagatgtgttactgttacacagaagtgtgtga 2985
Db 3001 GTTTCGTATACAGCTGTGTGCTTTTGATGTGTGTACTGTACACAGAGAGTGTGCA 3060
QY 2986 ctgaggtctgcgtgtgtgtccgatatgaaaaacctgtgtagccctgcgagttgaactgtct 3045
Db 3061 CTGAGGCTCTGCTGTGTGCTGTATGGAACCAACCTGTAGCCCTCGAGTTAAGTACTGCT 3120
QY 3046 tccatttactgtttagcgtgtgaatttctcccatgtgaatgtlaagtaaaactgaagtgt 3105
Db 3121 TCCATTCATTGTTTACGCTGGAATTTTCTCCCATGGAATGTAAGTAAACCTTAAGTGT 3180
QY 3106 ttgtcatcaataaatgtgtaataactaaaaaa 3142
Db 3181 TTGTCAATCAATAATGGTAATACTAAAAA 3217
RESULT 3
AK025519
LOCUS AK025519 3194 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21866 fis, clone HEP02379, highly similar to
AF061938 Homo sapiens stauflen protein (STAU) mRNA.
ACCESSION AK025519
VERSION AK025519.1 GI:10438059
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:Hepg2 cDNA to mRNA, clone_11b:HEP
clone:HEP02379.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

Qy	3121	ggtataactataaaaaaaa	3142
Db	3121	GGTATACTATAAAAAAAA	3142

RESULT	2
AF061938	
LOCUS	
DEFINITION	AF061938 3217 bp mRNA PRI 04-MAR-1999 Homo sapiens stauken protein (STAU) mRNA, alternatively spliced, complete cds.
ACCESSION	AF061938
VERSION	AF061938.1 GI:4335946
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 3217) Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L. Mammalian stauken is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL	99147057
MEDLINE	2 (bases 1 to 3217)
REFERENCE	DesGroselliers,L., Wickham,L. and Luo,M.
AUTHORS	Title Submission
TITLE	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
JOURNAL	

FEATURES	Location/Qualifiers
source	1. .3217

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. .3217
gene

```

```

/ gene="STAU"
363. .1853
CDS

```

```
/gene="STAU"  
/function="RNA binding protein"  
/note="alternatively spliced transcript"
```

```

/codon_start=1
/product="stauflen protein"
/protein_id="AAD17530.1"
/db_xref="GI:4335947"

```

/translation="MKLGKKKPMYPVDSRMOSTYNNMRGGAYPRRYEYPPVPL
 LYQVELSVGGQGFNGKGTROAKHDAAKALRILQNEPLPERLEVNGRESEENLNK
 SEISQVEIALKRLPLVNEVAREVAPRRIKKTPIKVPQTSVGGOGINPISRLAQIQAK
 ALAYLEELKLPPLPRAVERVMOVKYGHHTAEGTGNKKYAKRNAEENMLEIGFVPP
 EKEPEYTLITLTERGPRRREVMQVYGHHTAEGTGNKKYAKRNAEENMLEIGFVPP
 OROPTKPAKSEEXTPIKKPGDGRKATFEFPGSGDENGTSNKEDEFRMPLSHQOLPA
 GLIPMPEVAQAVASQGHHTKDETRAPNPAKATVTAMIARELVIGTSPAEITLKK
 NNISSGVHPGLTRPSEQLDYLISRYQGFQVEYKDFPKNNKNEFVSLINCSSOPPLIS
 NGTIGDVESCHDMAINIKTISLIDOSTEMPTGKPMASVCGRC"

BASE COUNT	ORIGIN
889 a	796 c 742 g 790 t

Query Match	97.2%;	Score 3053.8;	DB 10;	Length 3217;
Best Local Similarity	97.6%;	Pred. No. 0;		
Matches 3140; Conservative	0;	Mismatches 2;	Indels 75;	Gaps 1;

QY 1 acttccctgcgggactgacggcgccctctcagcgttgcgcgcgcgtcgcgcgt 60
|||||
Db 1 ACTTCCTGCCGGGCTGCGGGCGCCCTAGCGGCTCTCAGCGTTTGCGGGCGGCGCTGCGCGT 60

QY 61 ctctctgcgcctccgccttcccttgcacgcgcctcccccccgccgcgcgcgcgcctc 120
|||||
Db 61 CTCTCTCGGCTCCCGCTTCTCTTGACCGCCTCCCCCCCCCGCGCGCGCCGCGCTC 120

QY 121 ctccacggcgaactccgcctcttcccttcccttgcctcccttcttcccttcttccctt 180
|||||
Db 121 CTCACAGGCGCACTCCGCGCTCTCCCTCTGTCGTCCTCTTCCCTCTCCCTTTTTCCTT 180

[illegible]

QY 241 agcagcc----- 247

Db 241 AGCAGCCAGAGTTTATTAAACCACTTAACTCTCAGAACTGAACAAAGAACAACATTTGTTCC 300

OY 248 -----agaaagcataaccctactgtagaactaatgtcactgt 285

Db 301 TGGACGCCCTCTTTTAAAGAAAGCATACCCCTACTGTGAACTAATGCAC1GT 360

QY 286 gcatgaacttggaaaaaaccaatgtatagaagcctgttgaccctactctcgatgcagc 345

Db 361 GCATGAACCTGGAAAAAACCAATGTATTAAGCCTGTTGACCCCTTACTCTCGGATGCAGT 420

Qy 346 ccacctataactcaacatgagagagtgcttaccctcccgaggtactttaccacatttc 405
|||||

Db 421 CCACCTATACTACACATGAGAGAGTCTTATCCCCGAGTACTTTTACCCATTTC 480

[illegible]

Ov
465 aaataaaacaaagacaactatccaaaacacatctactaccsaagaattaaagatcctac 525

D0
481 CAGTLCACCTTTACTTTATCAAGTGAACTTC TGTGTAAGCAGCGAATTATAAGCA 549

Db 541 AAGGAAGACAAAGGCTGGCAACACGATGCTGCTGCCAAAGCGTTAGGATCCTGC 600

526 agaattgagcccttcacagagagagcttggaattgaatygaaagagaatccgaagaagaatac 585

Db 601 AGATGAGCCCCCTGCCAGAGAGGCTGCAGCTGATGGAAGAGATCCGAGAGAGAAATC 660

586 tcaataatctgaaataagtcagtggttgagattgcacttaacggaactgtcctgtga 645

Db 661 TCATAAATCTGAAATAGTCAAGTGTTTGAGATTGCACCTTAACCGAAGCTTGCCCTGTGA 720

QY 646 atttcgaggtgccccgggagagtggcccaaccacatgaagaacttgtgaccaagttc 705

Db 721 ATTTCGAGGTGGCCCGGAGAGTGGCCACCCACATGAAGAACCTTGTGACCAAGGTTT 780

QY 706 cgglttgggagtttctgtggggaagtgtgaagggaaaaaacgaagaagatttcaagaaaaaatc 765
|||||

Db 781 CGGTGGGAGCTTCTGGGGGAAGTGTAAGCGAATAAGCAGAGAAGTTCAAGAAAAATG 840

[illegible]

826 taagacctaagatacaaaaaataaaaaacacaaaccccatatcaaacccacacagacaaaccccaaat 885
DU 841 ccgccataagcctgatttcttgaagagcctgaagagaaataccgcgccctctgcttgcagctgaacgag 500

Db 901 TAAAGCTAGAAATCAAAAAGAAAACAAACCCATAGTCAAGCCACAGACACAAAGCCAGAAAT 960

QY 886 atggccagggatcaatccgattagccgactyggccagatccagcagycacaaaaagaga 945

Db 961 ATGGCCAGGGGATCATCCGATTAGCCGACTGGCCAGATCCAGCAGGCAAAAAAGAGA 1022

946 aggaagccaggtacacgcctcctcacagagcgaagcctccgcgcgcgcagggagtttgtga 1000

Db 1021 AGGAGCCAGAGTACACGCTCCTCACAGAGCGAGGCCCTCCCGCGCCGACAGGGAGTTGTGA 108

QY 1006 tgcagtgaggttcggaaccacactgcagagaaggaacgycaccaacaagaagtgcga 106

Db 1081 TGCAGGTGAAGGTTGGAAACACACACTGCAGAGGAACGGGCACCAACAAGAAGGTGGCCA 114

QY 1066 agcgcaatgcagccgagaaacatagtctgtgagatccttgtttlcaaaagttccgcgacgcgcagc 112

Db 1141 AGCGCAATGCAGCCSAGAACATGCTGGAGATCCTTGGTTTCAAGTTCGCCGAGCGGCAGC 120

1201 CCAACCAACCCCGGACTCCAAAGTCCAGAGGAGAAAGACATCCCATAAAGAAACCCAGGGGATGGAA 126
 1126 GCAACCAACCCCGGACTCCAAAGTCCAGAGGAGAAAGACATCCCATAAAGAAACCCAGGGGATGGAA 1181

C

Db 901 ATCCGATTAGCCGACTGGCCACGATCCAGCAGCAAAAAAGAGAAGAGCCACAGTACA 960
QY 961 cgcctctcacagagcgagggcctcccgcgccgcaaggagttgtgatgcaggtgaagtgtg 1020
Db 961 CGCTCCCTCACAGAGCGAGGCCCTCCGCCGCCGACGAGGAGTTGTGATGCAGGTGAAGTTG 1020
QY 1021 gaaaccacactgcagaaggaaggggcaaccaagaagtggtccaagcgcaatgcagccg 1080
Db 1021 GAAACCCACACTGCAGAAGGAAGGAGGGGCACCAACAAGAAGCTGGCCAAAGCCAAATGCAGCCG 1080
QY 1081 agaacatgctygagatccttggtttcaaagtcgccgcaaggcagcccaaccaaccgcac 1140
Db 1081 AGAACATGCTGGAGATCCTGGTTTCAAAGTCCCGCAGCGGCAGCCCAACCAACCCGCAC 1140
QY 1141 tcaagtcagaggagaagacacccataaagaaccaaaggagatggaagaaagttaaccttt 1200
Db 1141 TCAAGTCAGAGGAGAAGACACCACATAAAGAAACAGGGGATGGAAGAAAAAGTAACCTTTT 1200
QY 1201 ttgaaccggtctgygagatgnaaaatgggactagtaataaaggagatgagttcaagatgc 1260
Db 1201 TTGAACCTGGCTCTGGGATGAAATGGGACTAGTAATTAAGAGGATGAGTTCAGGATGC 1260
QY 1261 cttatctaagtcatcagcagctgcctgtcgtgaaattcttcccatggtgcgccgagtcgcc 1320
Db 1261 CTTATCTAAGTCATCAGCAGCTGCCTGCTGGAATTCCTCCATGGTGCCCGAGGTGCCCC 1320
QY 1321 aggcctgtaggagttagtcagaagacatcacaccaagaattttaaccaaggcagctccgaatc 1380
Db 1321 AGGCTGTAGGAGTTAGTCAAGGACATCACACCAAAAGATTTTACCAGGGCAGCTCCGAATC 1380
QY 1381 ctgccaaggcacagtgtaactgcacatgataggccgaagtgltgtatggygcacctgcgc 1440
Db 1381 CTGCCAAGGCCACGGTAGCTGCACTGATAGCCCGAGAGTGTGTATGGGGCACCCTGCGC 1440
QY 1441 ccacagccgagaccattttaagaataacatctcttcagggcaacgtaccatcgacctc 1500
Db 1441 CCACAGCCGAGACCATTTTAAAGAAATAACATCTCTCAGGCCACGTACCCTCATGGAACCTC 1500
QY 1501 tcacgagaccctctgagcaactggaactatcttccagagtcaggagattccaggttgaat 1560
Db 1501 TCACGAGACCCTCTGAGCAACTGCACTATCTTCCAGAGTCCAGGATTCAGGTTGAAT 1560
QY 1561 acaagacttccccaaaaaacaagaagaatltgtatctcttatacaattgtctctc 1620
Db 1561 ACAAGACTTCCCCAAAAACAACAAGAAGCAATTTGTATCTCTTATCAATTTGCTCTCTC 1620
QY 1621 agccacctctgatcagccatggtatcggcaagatgtggaagtcctcgcatgatagtctg 1680
Db 1621 AGCCACTCTGATCAGCCCATGGTATCGCCAAGATGTGAGTCTCGCCATGATATGGCTG 1680
QY 1681 cgctgaacatcttaaagttgtctgtcgtgagttggaaccaaaagtlacagagatgccaagaa 1740
Db 1681 CGCTGAACATCTTAAAGTTGCTGCTGAGTTGGACCACAAAAGTACAGAGATGCCAAAGAA 1740
QY 1741 caggaacggaaccaatgtctgtgtgtggyaggtgtcgaacctttctcgcatgaacct 1800
Db 1741 CAGGAACGGACCAATGTCTGTGTGTGGAGGTGCTGAACCTTTTCTGCCCATGAACCAT 1800
QY 1801 tataaatcccaacatatatactgaaataactgaaactgcttggaaatttgaaatttct 1860
Db 1801 TATAAATCCCAACATATATACGAAAAATACTGAAACTGCTTTGAAAAATTTGGAATTTCT 1860
QY 1861 gatacctcagtggtgagagaaacacaggtgtggttaaagatgtggycagcagcaggyaaga 1920
Db 1861 GATACCTTCAGTGGGCCGAGAGACACGGGTGGTAAGGATGTGGGCAGCAGCAGGGAAGA 1920
QY 1921 caacagaaacacaaggaagcggtgtgtgycggctgagctgtgtggygttgtgtgatg 1980
Db 1921 CAACAGAAACACAAGAGGCGGCTGTGGCCGGCTGGACTGTGCTGGGGTTGTGTGATG 1980
QY 1981 gccactcggtgacctggtgcgttcctacgcaatagcagctgcctgtgtggyaagaaggctg 2040
Db 1981 ggcctggaatttttctccccaatgtaagtaaaactttaaagttgttgcatacaataat 3120

Db 1981 GCCACTCGGTGACCTGGCGGTCCTCTACGCAATAGCAGCTGCTGTGGGGAAGAGGCTG 2040
QY 2041 cccagcagctggtlctcccggyaacaacagagatccacaacctgygcaactcgtgtlt 2100
Db 2041 CCCAGCCAGCTGGTCTCCCGGGACACCAGAGATCCACACCCTGGGCACTCCGTGTTT 2100
QY 2101 ggtcttttttccctctgtgtgaagaagaagaacgycacagcccttctcaagctggtc 2160
Db 2101 GGTCTTTTTTTCCCTGTGTGAAGAAGAAACGGCAGCAGCCCTTCTCAAGCTGGCTCA 2160
QY 2161 ctcaagacacatggygacaacacctgycacagccatgccaagagagagcgcttgcggccc 2220
Db 2161 CTCAGACACATTTGGGACAAACCCCTGGACAGCCATGCCAGAGAGAGGCCCTTGAACGGCCC 2220
QY 2221 cagagctaaaagcacacagaganaaatacaatgcttctactcaagcgtgaaccaatttct 2280
Db 2221 CAGAGCTAAAGCACACAGAGANAATCAATGCTTCTTACTCAGCGTGACCCCACTTTTCT 2280
QY 2281 agtgcacaagcccccaaccctcctgcagtaaccacaaccaataccaactgcttctctc 2340
Db 2281 AGTGTCCACGGGCCCCACACCTCTCGAGTACCACACCATCACCACTGCTTCTCTTC 2340
QY 2341 caacagtgatctgtatcttgaattcaatatttcttltgaltgatagcacatataa 2400
Db 2341 CAACAGTATCTGTATCTTAGTTCAATTATTTCTTTGATTGATATGACACTATATAA 2400
QY 2401 aatttcaattgagaatttctcaattgtlatctagttlaatatgacagatttggaaattgt 2460
Db 2401 AATTTTCAATTGAGAAATTTCTCAATTGTATCTAGTTAAATAGCACAGTTTGAACCTTGT 2460
QY 2461 ctgagactgacttlatcaataatctaaoccgacaaagatcatatccatgtgtatgtgtta 2520
Db 2461 CTGAGACTGACTTTATCAATATCTAACCAGACAAGATCATATCCATGTGTATGTGTTA 2520
QY 2521 gacatttlatlctcatgtactaaaccagagacagtttcaagtgtaagaaattgtgtccctc 2580
Db 2521 GACATTTTATTTTCAATTGACTTAACCCAGGACAGTTCAGTGATGCCAAATTTGTGTGCCCTC 2580
QY 2581 tgttcagctgaaaacagtcctcgtgacttccaacaaacctgaaataagtcctcccaagttgta 2640
Db 2581 TGGTTCACTGAAAACAGTCCCTGACTTTCAAAAACCTTGAATAAGTCTCCACAGTTGTA 2640
QY 2641 taaattgacaattlaggaattttaaactttagatgatacttggltccatlttatttc 2700
Db 2641 TAAATTGACAATTTAGGAATTTTAAACTTTAGATGATCATTTGGTTCATTTTATTTTC 2700
QY 2701 attttatttttgttaatgcaaacagagacttaaatgaacttgaatctcgttlttaaat 2760
Db 2701 ATTTTATTTTGTGTTAATGCAACAGGACTTAAATGAACCTTGATCTCTGTTTAAAGAT 2760
QY 2761 tattaaaaaacattgtgtatctcatatacatatgycctctgagagacttagcttcaactaact 2820
Db 2761 TATTTAAAAACATTTGTGTATCTATACATATGGCTCTTGAGGACTTAGCTTTCACACT 2820
QY 2821 acaagatagatctccatgtagtccatatataaacctgcagagatgatttccaaggtgctg 2880
Db 2821 ACAGGATATGATCTCCATGTAGTCCATATAAACCTGCAGAGTGATTTTCCAGAGTGCCTG 2880
QY 2881 atactgttaattacatctccattagggtgaaagaatgacctacglttctgtatacagc 2940
Db 2881 ATACTGTTAATTACATCTCCATTAGGGCTGAAAAGAAATGACCCTACGTTTCTGTATACAGC 2940
QY 2941 tgtgtgcttttgatgtgtgttactgttacacagaagtgltgcactgagcgctcgtgtg 3000
Db 2941 TGTGTGCTTTTGTGATGTTGTGTACTGTACACAGAGAGTGTGCACTGAGGCTCTGCGTG 3000
QY 3001 tgytccgfatgynaacacctggtagccctgcygagttgaagtaactgcttccatctcatgtta 3060
Db 3001 TGGTCCGTATGGAACACCTGGTAGCCCTGCGAGTTAAGTACTGCTTCCATTTGTTT 3060
QY 3061 cgcctggaatttttctccccaatggaatgtgaagtaaaacttaaagttgttgcatacaataat 3120
Db 3061 CGCTGGAATTTTCTCCCCATGGAATGTAAGTAAACTTAAAGTGTGTGTATCAATAAAT 3120

22	150.8	4.8	160979	52	AC055879	AC055879 Homo sapi
23	149.6	4.8	3173	34	AK001576	AK001576 Homo sapi
24	149	4.7	332811	72	AL390198	AL390198 Homo sapi
25	133.4	4.2	194407	54	AC068845	AC068845 Homo sapi
26	125.2	4.0	2008	34	AK023314	AK023314 Homo sapi
27	119	3.8	2847	69	HSRNATFTI	X83973 H. sapiens m
28	110.2	3.5	5237	53	DROSTAUFEN	M69111 D. melanogas
29	110.2	3.5	61267	28	AC004336	AC004336 Drosophil
30	110.2	3.5	120625	38	AC017563	AC017563 Drosophil
31	110.2	3.5	267271	28	AE003800	AE003800 Drosophil
32	105.8	3.4	8409	29	AF225924	AF225924 Drosophil
33	103	3.3	8002	29	AF225925	AF225925 Musca dom
34	85.4	2.7	171591	50	AC027018	AC027018 Homo sapi
35	70.8	2.3	7218	81	I66494	I66494 Sequence 14
36	68	2.2	7218	81	I66494	I66494 Sequence 14
37	65.4	2.1	60009	48	AC023436	AC023436 Homo sapi
38	63	2.0	180968	62	AC078996	AC078996 Mus muscu
39	63	2.0	258550	35	AC010758	AC010758 Homo sapi
40	62.6	2.0	42099	29	CELF55A4	U67949 Caenorhabdi
41	62.6	2.0	154746	80	HSV2HG52	Z86099 Herpes simp
42	62.6	2.0	154746	80	HSV2HG52	Z86099 Herpes simp
43	62.4	2.0	141892	47	AC023197	AC023197 Mus muscu
44	62.2	2.0	89669	69	HSJ340H11	AL109755 Human DNA
45	61.4	2.0	77127	36	AC011801	AC011801 Homo sapi

ALIGNMENTS

RESULT	1
LOCUS	AF061940
DEFINITION	AF061940 3142 bp mRNA PRI 04-MAR-1999
ACCESSION	Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced, complete cds.
VERSION	AF061940
KEYWORDS	AF061940.1 GI:4335950
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 3142)
TITLE	Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L. Mammalian stauflen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE	99147057
REFERENCE	2 (bases 1 to 3142)
AUTHORS	DesGroselliers,L., Wickham,L. and Luo,M.
TITLE	Direct Submission
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES	Location/Qualifiers
source	1..3142
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q13.1"
	1..3142
	/gene="STAU"
	288..1778
	/gene="STAU"
	/function="RNA binding protein"
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="stauflen protein"
	/protein_id="AAD17532.1"
	/db_xref="GI:4335951"
	/translation="MTLGGKPPMKPVDPYSRMOSTYNYNMRGGAYPRRYEYPFPVPL LYQVELSVGGQQFNGKGKTROAKHDAAKALRLIQNEPLPERLEVNGRSEENLNK SEISQVEIETALKRNLPVNFEEVARESGPPhMKNVPTKSVEGEFEGEGSKISKNN AIAVLEELKKLLPPLPAVERVKPRIKRTTKPIVKQTSPEYGGINPISRLAQIOAKK EKPEPYTLITTEGLPRRRREFVMQVKNHTAEGTGNTKKVAKRNAENMLEILGEKVP
CDS	
gene	

BASE COUNT		862 a	778 c	733 g	769 t	ORQRPKPAALKSEKTPPIKKPGDGKRVTFEFPGSGDENGTSNKEDFMRPYLSHQQLPA GILPMVDEVAQAVGSOGHDTKDFRAAPNPAKATVTAMIARELYGTSPTAETILK NNISSGHVPHGPLTRPSEQLDLSRYQFQVEYKDFPKNNKNEFVSLINCSSQPLIS HGIGKDYESCHDMAALNIILKLSELDQOSTEMPRGTNGNPMsvCGRC"
Query Match		99.9%;	Score 3138.8;	DB 10;	Length 3142;	
Best Local Similarity		99.9%;	Pred. No. 0;			
Matches 3140;		Conservative	0;	Mismatches	2;	Indels 0;
QY	1	acttcctgcgagctgcgagcgccctgagcgctctcaagcgtttgcgcgcgctgcgct	60			
Db	1	ACTTCCTGCCGGGCTGCCGGGGCCCTGAGCGCTCTTCACGCTTTGCCGGGGCTGCCGCT	60			
QY	61	ctctctcgctcccgctctctcttgaacgcctcccccccccgccgcgcgccgctc	120			
Db	61	CTCTCTCGGCTCCCGCTCTCTTTCCTTTGACCGCCTCCCCCCCCGGCCGGCGCCGCTC	120			
QY	121	ctccacgycgaactccgcctcttccctccctcgttccctctctcctcttctt	180			
Db	121	CTCCACGGCCACCTCGGCTCTCTCCCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCT	180			
QY	181	cttccttccccctctcgccgcacccgcccagagaccgcccggggagagctcgagc	240			
Db	181	CTTCCTTCCCTCTCTCGCGGCCACCGCCAGAGACCGCGGCGGGGAGACGATCCGGAGC	240			
QY	241	agcagccagaagaacataaccctactgttagaactaaatgcactgtgcatgaaacttgaa	300			
Db	241	AGCAGCCAGAAGCATATACCCCTACTGTAGAACTAAATGCACGTGCATGAACCTTGAA	300			
QY	301	aaaaaccaatgtataaagcctgttgaccccttactctcgatgcagtcacactataactaca	360			
Db	301	AAAAACCAATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATACTACA	360			
QY	361	acatgagagaggtgcttatacccccgaggtacttllacccattlccagttccaccttac	420			
Db	361	ACATGAGAGGAGGTGCTTATTCCTCCCGAGGTACTTTTACCCATTTCACGTTCCACCTTAC	420			
QY	421	ttatcaagtgyaaccttctgtgagagacagcaacttlaatgycaaagaaagacaagac	480			
Db	421	TTATCAAGTGGAACTTTCTGTGGAGGACAGCAATTTAATGGCAAAAGAAAGACAAGAC	480			
QY	481	agcgtgcgaacacacgatgctgcgtgccaaagcgtltgagatcctgcagaatgagccctgc	540			
Db	481	AGGCTGCGAAGAACACGATGCTGCTGCCAAAGCGTTGAGGATTCCTGCAGATGAGCCCTGC	540			
QY	541	cagagagcgtgaggtgaaatgyaagagaatccgaagaagaanaatctcaataatctgaaa	600			
Db	541	CAGAGAGGCTGGAGGTGATGGAAGAGAAATCCGAAGAGAAAATCTCAATAATCTGAAA	600			
QY	601	taagtcaagtgtttgagatgtcaacttaaacggaactgcctgtgaaattcgaggtgccc	660			
Db	601	TAAGTCAAGTGTTTGAGATGCACTTAAACGGAACCTTGCTGTGAATTTGAGAGTGGCC	660			
QY	661	gggagagtgccccacccccacatgaagaactltgtgaccaaggtttcgtltgggagtttg	720			
Db	661	GCGAGAGTGGCCACCCACATGAAGAACTTTGTGACCAAGGTTTCGTTGGGAGTTTG	720			
QY	721	tgggggaagtgtaagggaaagcaagaagatttcaagaagaaatgcccgcatagtgttc	780			
Db	721	TGGGGGAAGGTGAAGGGAAGCAAGAAGATTTCAAAGAAATAATGCCCGCATAGCTGTTC	780			
QY	781	ttgagagctgaagaagttaacgcgccctgcctgcagtltgaacgagtaagcctagaatca	840			
Db	781	TTGAGGAGCTGAAGAAGTTACCGCCCTTGCTGCAGTTGAACGAGTAAAGCCTAGATCA	840			
QY	841	aaaaaagaaacaaaccccatagtcaagccacagacaagaagcccagaatatatgcccaggygatca	900			
Db	841	AAAAAGAAACAAAAACCATATAGTCAAGCCACAGACAAGCCACAGATATATGGCCAGGGGATCA	900			
QY	901	atccgattagccgactgccccagatccagcagggcaaaaaaggaagagccagagataca	960			

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2001, 10:59:15 ; Search time 24649.1 Seconds
(without alignments)
652.356 Million cell updates/sec

Title: US-09-316-048-1
Perfect score: 3142
Sequence: 1 acttcctgcggcgctgcggg.....taataactaaaaaaaaaaaaa 3142

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: GenEmbl:*
- 2: gb_bal:*
- 3: gb_ba2:*
- 4: gb_com:*
- 5: gb_ov:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: em_fun:*
- 13: em_hum1:*
- 14: em_hum2:*
- 15: em_in:*
- 16: em_om:*
- 17: em_or:*
- 18: em_ov:*
- 19: em_pat:*
- 20: em_ph:*
- 21: em_pl:*
- 22: em_ro:*
- 23: em_sts:*
- 24: em_sy:*
- 25: em_un:*
- 26: em_vl:*
- 27: gb_htg1:*
- 28: gb_htg2:*
- 29: gb_in1:*
- 30: gb_in2:*
- 31: em_bal:*
- 32: em_ba2:*
- 33: em_hum3:*
- 34: em_hum4:*
- 35: gb_pr4:*
- 36: gb_htg3:*
- 37: gb_htg4:*
- 38: gb_htg5:*
- 39: gb_htg6:*
- 40: gb_htg7:*
- 41: em_htg1:*
- 42: em_htg2:*
- 43: em_htg3:*
- 44: em_hum5:*

- 44: gb_pl3:*
- 45: gb_pr5:*
- 46: gb_htg8:*
- 47: gb_htg9:*
- 48: gb_htg10:*
- 49: gb_htg11:*
- 50: gb_htg12:*
- 51: gb_htg13:*
- 52: gb_htg14:*
- 53: gb_in3:*
- 54: gb_htg15:*
- 55: gb_htg16:*
- 56: gb_htg17:*
- 57: em_htg4:*
- 58: em_htg5:*
- 59: em_htg6:*
- 60: em_htg7:*
- 61: em_hum6:*
- 62: gb_htg18:*
- 63: gb_htg19:*
- 64: gb_ba3:*
- 65: em_htg8:*
- 66: em_htg9:*
- 67: em_htg10:*
- 68: gb_pr6:*
- 69: gb_pr7:*
- 70: gb_htg20:*
- 71: gb_htg21:*
- 72: gb_htg22:*
- 73: gb_htg23:*
- 74: gb_ro:*
- 75: gb_sts1:*
- 76: gb_sts2:*
- 77: gb_sy:*
- 78: gb_un:*
- 79: gb_vil:*
- 80: gb_vil2:*
- 81: gb_pat1:*
- 82: gb_pat2:*
- 83: em_htg0:*
- 84: gb_htg24:*
- 85: gb_pr8:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query		DB	ID	Description
No.	Score	Match	Length			
1	3138.8	99.9	3142	10	AF061940	AF061940 Homo sapi
2	3053.8	97.2	3217	10	AF061938	AF061938 Homo sapi
3	2945.2	93.7	3194	34	AK025519	AK025519 Homo sapi
4	2921.8	93.0	3349	10	AF061941	AF061941 Homo sapi
5	2896	92.2	3506	10	AF061939	AF061939 Homo sapi
6	2799.4	89.1	3066	68	HSAL32258	AJ132258 Homo sapi
7	1357.4	43.2	113836	34	AL133174	AL133174 Human DNA
8	1356.2	43.2	2860	74	AF061942	AF061942 Mus muscu
9	1354.6	43.1	3041	74	AF290989	AF290989 Rattus no
10	1326.6	42.2	3059	74	AF227200	AF227200 Rattus no
11	705	22.4	157803	35	AC011492	AC011492 Homo sapi
12	553.2	17.6	194407	54	AC068845	AC068845 Homo sapi
13	383.2	12.2	192597	70	AL158850	AL158850 Homo sapi
14	380.2	12.1	385	75	G30939	G30939 DRESSO Huma
15	287	9.1	360	74	RNO010200	AJ010200 Rattus no
16	283.6	9.0	225016	52	AC063967	AC063967 Mus muscu
17	278.2	8.9	2968	34	AK002152	AK002152 Homo sapi
18	278.2	8.9	4058	68	HSAG9062	Y19062 Homo sapien
19	220.2	7.0	263	75	G22107	G22107 human STS w
20	215	6.8	160979	52	AC055879	AC055879 Homo sapi
21	193.6	6.2	163253	68	HSB140E4	AL121877 Human DNA

QY 1705 agctgctgtctgagltggaaccaacagagacacagagatgccaagaacagaaatgaccag 1764
|||||
Db 1614 AGCTGCTGTCTGAGTTGGACCAACAGAGACAGAGATGCCAAGACAGAGAAACGACCAG 1673
QY 1765 ttccagcgtgcgagagtgctgaacctttctgcacaaacattataaaccacat 1824
|||||
Db 1674 TTTCAAGCGTGGGAGGCTGAACCTATCTCGCCATGAACCATATATAAACCCACAT 1733
QY 1825 atatactgaataactggaactgctttgaaaaatttggaatatctgataactccagtgg 1884
|||||
Db 1734 ATATACCGAAATACTGGAACACTGCTTTGAAAATTGGAAATTCTGATTAACCTCAGTGGG 1793
QY 1885 ccaagacatggtgat-aaaaatgtgccaagacgacaagaactcaggtgtagccc 1942
|||||
Db 1794 CCAAGACATGTTGATTAATAAATGTGGCAAGAAGAAAAGAACTGCAAGTGTGGCT 1853
QY 1943 tggttgtgctgcgctagtgatgatgtgtgtctctccatccacagacagaaccag 2002
|||||
Db 1854 TGGTTGGGCTG--GGCTGCTGGTGACGCTGTGCTCCGCCATCCA-CCAGAGACCAGCAG 1910
QY 2003 ccccaacgcctcagtgctgttttgcacgtgaca------gagagacagccaatt 2056
|||||
Db 1911 CTCCAGTGTCTCCAGCTTTTCTTTTGGCTCGTGACAGAGAAGAACGGAGAGCGCTGCCAACT 1970
QY 2057 ctcatgctgctctctcagataactttgaaaaaccccgacagccacacacagagagcccta 2116
|||||
Db 1971 CTCACGCTGGCTTCTTCAGATCTTTTGAAAACCGGGACAGCCACACAGAGGCCCTTA 2030
QY 2117 tagcgcccgagagctaaacgagacagagaaagccagtgcttctctactgcacatgact 2176
|||||
Db 2031 CAGCGGCGCCGGAGCTAAACGCACACAGAGAAA-ACCAGTGTCTCT-ACATGACT 2084
QY 2177 gactcagctccgccaacgtagca-ccactgttaaccactgcttctctcag-2228
|||||
Db 2085 GACGCAACTCTGCCACACAGTACACCCACCGTAACCACTGCTTCTCTCAACAGTGATG 2144
QY 2229 -----ttcatttttcccttgatgatacaacacta--taatttcattca 2274
|||||
Db 2145 GATGCCCTGTTAGTTCATATATCCCTTTGGTTGAGATGACACTATGTAATTTCATTTC 2204
QY 2275 gtcccttagtcgtgtctactacctagcagttagaactgt-----2316
|||||
Db 2205 GTTCTCAGTTGATCTACTTACCTAGCAGTTTGAACACTAGTCTGAGACAGATGCTATC 2264
QY 2317 cagtcatgtlaactgcaagatcacagcccggttggtg-gcattctgtgctctgct 2374
|||||
Db 2265 CAGCTATGTAAGTGGCAAGACACAGACCCCTGTTGGTGATGTCNTTCTGTGCTCTGGCT 2324
QY 2375 tggctgaacagtlctggaattacacacagaatcctgactccctgccccctgtataaatt 2434
|||||
Db 2325 CGGCTGGCAGTCTGTGACTTANCCACAAGAACTTGACTCCCAACCCCTTGATTAAT 2384
QY 2435 ggaacagcttaggaactttaaaactt--agatcaaaagatatgtccttttaacttlat 2491
|||||
Db 2385 GGACAGCTTAGGACTTTTAACTTTAGAGATCAAAAGATTGGTCTTTTAACTTTAT 2444
QY 2492 tttaaggaagcagactttaaaatgagccctgaccttaaccattataacaga-----2543
|||||
Db 2445 TTTTAAAGCAGCATACTATAAAGTGAG-CCTGACCTTTAACAGTTATACAGATTTGTCAC 2503
QY 2544 -----attgtcaaaaggagtgcttctctgagaggta---2575
|||||
Db 2504 GGGAGAGGTTGTACCCCATTTTAACAGATTGCCAGTGGGAGGTCACCTTGAGGGAGTATTG 2563
QY 2576 -----gcttttttaccacactacagagacatcctgtaggcccaagactacag 2628
|||||
Db 2564 CTCGAAGGTGGCTTTTCTACACACTACAGGACATTAACCTGTGGGCCAGAAAGACTATGGG 2623
QY 2629 ctggtgtcccttagagggccaataacagtcacatccaacctctaagtcggggaagtgac 2688
|||||
Db 2624 CTGCTGTCCCTGGAGGGCCCAATACAGTCAATCCAGCCTCCACGTA-GGGAAGGGCGAC 2682
QY 2689 aggttctcgtgtgctg-----gtgtcacaggggcaaggcaggtcagctgctggggaag 2743

Db 2683 AGGTTCTCTGCTGTGCTGTGCTAGTGTGCACAGAGGACGAGAAGTCTGCCGCGCTGTGG 2742
QY 2744 agcatgtgtcctcctagtcagcc---ctgcttccactctgtgttagctggaacctcc 2800
|||||
Db 2743 AGCAATTGTGGCTCCGAGTGCACCCCTGCTGCTGCCACTCGGGGTTAGCTGG-ACCTTTT 2801
QY 2801 cactcatggaataaagtaaacctcaacttcttgcacccaataatgtaataactaa 2857
|||||
Db 2802 CACCCATGGAATATAAGAACTTA--GTCTTTGTCAACCAATAAATGAGAAATACTAA 2856

RESULT 3
AF227200
LOCUS AF227200 3059 bp mRNA ROD 09-AUG-2000
DEFINITION Rattus norvegicus staufen isoform Stau+I6 (Stau) mRNA, complete cds, alternatively spliced.
ACCESSION AF227200
VERSION AF227200.1 GI:9754865
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3059)
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desgroselliers, L., Kuhl, D., Richter, D., and Kindler, S.
TITLE Two rat brain staufen isoforms differentially bind RNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3059)
AUTHORS Monshausen, M. and Kindler, S.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) University of Hamburg, Institute for Cell Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg 20246, Germany
FEATURES
source location/Qualifiers
1. 3059
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1. 3059
/gene="Stau"
227. 1714
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau+I6"
/codon_start=1
/product="staufen isoform Stau+I6"
/protein_id="AAF98119.1"
/db_xref="GI:9754866"
/translation="MYKVPDPSHRMSTYSYGMRGAYPPRYFPFPVPLLYOVELS
VGGOQFNGKGMKRPVKHDATARALRLOSEPLPERLEVNGRESEENLNKSEISOVF
EIALKRNLPVNFESFPLTQVARESGPPHMKNEVTRVSVGEFVGEGSKSKISKRNA
RAVLEQLRLPLPAVERVKPRIKKSQPTCKLOTAPDYGGMNPISRLAOIOAKKE
KEPEYMLTERGLPRRREFVMQVKGHTAEGAGTNKKVARKNAENMLEILGFYVQ
AQPAKPAKSEKTPVKKPGDKRYTFEESP GDENGTSNKREDEFMRMPLYLSHOQIDAG
ILPMVPEVAQAVSVSGHHTKDFARAAPNPAKATVTAMIARELLYGTSPTAETIKS
NISSGHVPHGPRTRPSEQLYLISRAQGFQVEYKDFPKNNKNECVSLINCSSQPLVSH
GIGKDVESCHDMAALNLIKLSLSDQOSTEMPTGTNGPVSACGRC"
3043. 3048
polya_signal
/gene="Stau"

BASE COUNT 817 a 778 c 749 g 713 t 2 others
ORIGIN
Query Match 69.5%; Score 1985.4; DB 74; Length 3059;
Best Local Similarity 86.8%; Pred. No. 0;
Matches 2513; Conservative 0; Mismatches 233; Indels 149; Gaps 24;
QY 91 gcgtccccccttctctgagcccgccgctgagcgagcccgcccttgcctccgcaactcgc 150

[illegible][illegible]

QY	2259	ta-taatttcatttcagttccttagtcgtgtctacttaccctagcagtttagaaact--	2314
Db	2205	TATGTAATTTTCATTTCAGTTTCTCAGTTGCATCTACTTACCCTAGCAGTTAGAAACTAG	2264
QY	2315	-----gtcagtcagttaactgycgaagatcacagcccggttcggtg--g	2356
Db	2265	TCTGAGACAGATGCTATCCAGCATTGTAACTGGCAAAAGACCACAGCCCTGTTGGTGATG	2324
QY	2357	cattctgtgcctctgcttgcttgcaacagtlcttgyaatlacccaccagaatcccttgactcc	2416
Db	2325	CNTTCTGTGCCTCTGTGGCTCGGCTGGCAGCTTCTGGACTTANCCACAGAATCTTGACTCC	2384
QY	2417	ctgcccccttgataaattgagacagcttaggaactttaaactt--agatcaaaaagatat	2473
Db	2385	CAACCCCTTGATATAAATTGGACAGCTTAGGACTTTTAAACTTTAGGAGATCAAAAGATT	2444
QY	2474	ggtccttlttaacttatttlttaagagagcagactttaaagtgccttgacccttaacca	2533
Db	2445	GGTCCCTTTTAACTTATTTTAAAGCAGCATACTTAAAGTGAG-CCTGACCTTTACCAG	2503
QY	2534	ttataacaga-----attgtcaaaaggagtg	2560
Db	2504	TTATAACAGATTGTCAACGGGAGAGGTTGTCACCCATTTTAAACAGATTGCCAGTGGAGG	2563
QY	2561	ttcttgaggagtgta-----gcttlttlttaccacactacaggaacttacctgta	2610
Db	2564	TCACTTGAGGAGTATTTGCTCAAGGTGGCTTTTCTACACACTACAGACATTTACCTGTG	2623
QY	2611	ggccccaagaactacagagctgtgtcccttagagggcccaatacagtcacttcaacctct	2670
Db	2624	GGCCCAAGAAGACTATAGGCTGTGTCCTTGAGGGGCCCAATACAGTCAATCCCAAGCCTCC	2683
QY	2671	aagtcggggaaggtgacaggttccctgtgtgcty----gtgtgcacaggggcaggcagg	2725
Db	2684	ACGTA-GGGAAGGGCGACAGGTTTCCTGCTGTGCTAGTGTGCACAGAGCGAGAAGG	2742
QY	2726	tcaagctgacctgggggaagagcattgtgtctcctagtcagcc--ctgcttccactcttg	2782
Db	2743	TCTGCCCGCCCTGTGTGGAGCATTTGTGCTCCGAGTGCACACCCCTGCTGCCCACTCGGG	2802
QY	2783	gttagcttgyaaccttcccactcatggaataataagtaactcaacttcttgttccaacaat	2842
Db	2803	GTTTAGCTGG-ACCTTTTCACCCCATGGAATATAAGGAAACTTA-GTCTTTGTACCACAAT	2859
QY	2843	aaatgtaataactaa	2857
Db	2860	AAATGAGAATACTAA	2874
RESULT	4		
LOCUS	AF061940	3142 bp	mRNA
DEFINITION			Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced, complete cds.
ACCESSION	AF061940		
VERSION	AF061940.1	GI:4335950	
KEYWORDS			
SOURCE			human.
ORGANISM			Homo sapiens
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS			1 (bases 1 to 3142)
TITLE			Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
JOURNAL			Mammalian stauflen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum
MEDLINE			Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
REFERENCE			99147057
AUTHORS			2 (bases 1 to 3142)
TITLE			DesGroselliers,L., Wickham,L. and Luo,M.
JOURNAL			Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
			Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada

```

FEATURES
    source
        1. .3142
            /location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="20"
            /map="20q13.1"
            1. .3142
            /gene="STAU"
            288. .1778
            /gene="STAU"
            /function="RNA binding protein"
            /note="alternatively spliced transcript"
            /codon_start=1
            /product="staufen protein"
            /protein_id="AADI7532.1"
            /db_xref="GI:4335951"
            /translation="MKLGKKPKPVDPYSRMQSTYYNNMRGAYPPRYEYFPFPVPL
            LYQVELSVGGQJFNGKGTROAKHDAAKALRLQNEPLPERLEVNGRSESEENLNK
            SEISQVFETALRKNLPVNFVARESGPPHMKNFVTKVSQVGEFGEEGSKKISKNA
            AIAVLEELKRLPLPAVERVKPRIKKTKPIVKPQTSPEYGQGINPISRLAQIOQAK
            EKEPEYTLTERGLPRRREEVMQKVGHTAEGTGNKKVAKRNAENMLEILGFQVP
            QROPTKPAKSEKTEPIKKPGDGRKVFEEBGSNGENGTSNKEDFRMPYLSHOOLPA
            GILPMVEVAQAVGSQGHHTKDEFTRAADNPAKATVTAMIAARELLYGTSPTAETILK
            NNISGHHVPHGLTRPSEQDLYLSRVQGFQVEYKDEFPKNNKNEFVSLINCSSQPLIS
            HGIGKDVESCHDMAALNLIKLSLSDQOSTEMPTGTNGPMSVCGRC"
BASE COUNT      862 a      778 c      733 g      769 t
ORIGIN

```

Query Match	47.48;	Score 1354.6;	DB 10;	Length 3142;
Best Local Similarity	84.68;	Pred. No. 0;		
Matches 1597; Conservative	0;	Mismatches 269;	Indels 21;	Gaps 6;

QY	23	ac	lccctc	gcagc	gtc	gcgg	gcccc	gagc	gcgtc	ctctc	aagc	gtttg	cgtc	gtc	gtc	gc	82
Db	1	ACTTCCTG	CGGGGCTG	CGGGCG	CGCTTG	AGCGCTCT	TCAGCGTTT	GGCGGC	-	-	GGCTG	CGC					58
QY	83	gtc	tgtg	tgc	gc	tc	ccccctc	ctctc	tgag	-----	ccccg	ccctg	cg	gc	ccgc		132
Db	59	GTC	TCTCTCG	GGCTCC	GGCTTCTTC	TGACCG	CCCTCCCC	CCCCCG	CGCGCG	CGCC							118
QY	133	ttc	gcctc	gcg	cactc	cg	ccctctctc	ctc	ctc	tgtg	tgc	gcccc	ttt	ctc	tc	gc	192
Db	119	TCCTTC	CACGGCCACT	CGGCTCTTC	CCCTTC	GTCCCTTC	CTCCTCTC	CTCCTCTC	CTCCTCTC	CTCCTCTC							178
QY	193	cact	tgtctc	tca	ctc	ctc	tcgc	gcgcg	cccc	aagacc	gcgcg	ccccg	gag	cag	ctc	tg	252
Db	179	TTC	TTCCCTTCC	CTCTCTCG	CGCCAC	CGCCAG	ACCGCG	CGGGG	GAGAG	-TCCGG							237
QY	253	gga	agcag	ccaga	aa	gtata	gctc	tca	ccatt	tgatc	taatg	caatg	caatg	tg	tg	gaa	312
Db	238	AGC	AGCAGCCAG	AAAGCAT	TAACCCCT	ACTGTAG	AACTAA	ATGC	ACTGTGC	ATGA	AACTTG						297
QY	313	aa	agaa	aa	ccaat	gtata	ga	gcc	cg	tga	ccctc	actc	ctc	gc	gt	ca	372
Db	298	GAA	AAAAAACCA	TGTAT	TAAGCC	TGTG	ACCCTT	ACTCTCG	GATGC	AGTCC	ACCTAT	ACT					357
QY	373	atg	catg	cgt	lga	aggtgc	ctat	cccc	ccag	ata	actt	tl	ta	cc	att	c	432
Db	358	ACA	ACATG	AGAG	AGGTG	CTTAT	CCCCCG	AGGTACTT	TATCC	CAATTTC	CAGTTCC	ACCTT					417
QY	433	tact	ctacca	agt	tg	agctc	tc	ccg	tgc	gcg	a	gcag	ca	gtt	ta	tg	492
Db	418	TACT	TTATCA	AGTGA	CTTCTCT	GTGG	AGGAC	AGCA	CAATT	TAA	TGGCA	AAAG	GAAG	ACAA			477
QY	493	ga	ccacc	cg	tga	aa	ca	cgat	gcccc	tc	gc	gt	gag	ga	ct	gc	552
Db	478	GAC	AGGCTG	CGA	AACG	ATGCTG	CTGCC	AAAG	CGTTG	AGGAT	CTCG	AGAT	GAG	CCCC			537
QY	553	tg	ccaga	aa	ag	gtc	tg	ag	gt	ta	at	g	ga	ga	ga	ga	612
Db	538	TGCC	CAGAG	AGGCTG	GAGGTGA	ATGGA	AGCA	ATCCGA	AGAG	AAATCT	CAAT	AAATCTG					597

QY	613	aaataagccaagtgtttgnaaattgcgctgaagcggnaattgcctgtgaattttgagttg	672
Db	598	AAATAAGTCAAGTGTtTGAGATTGCACCTTAACCGGAACCTTGCCCTGTGAATTCAGAGTGG	657
QY	673	cccggaagatlgccccacacatgaagaactttgtgaccaggtttcagttgggaat	732
Db	658	CCCCGAGAGTGGCCACCACCACATGAAGAACTTTGTGACCAAGGTTTCGGTTGGGAGT	717
QY	733	ttgtaggggaaggagaaggaaaagcaagaagatctccaagaagaatlgcgcgcaggtctg	792
Db	718	TTGTGGGGGAAGGTGAAGGGAAGCAAGAAGATTTCAAAAGAAAAATGCCGCCATAGCTG	777
QY	793	ttctgagcagccttagaggtgtccacccctccctgtgtgagcgagtgaagcccaaa	852
Db	778	TTCTTGAGGAGCTGAAGAGATTACCGCCCCCTGCCCTGTCACTTGAACGAGTAAGCCTTAGAA	837
QY	853	tcaagaagaaaagtcagcccacctgca-----agacagccccggatlatgcccgaagga	906
Db	838	TCAAAAAGAAAACAAACCCTAGTCAAGCCACAGACAGCCCAAGATATGGCCAGGGGA	897
QY	907	tgaatcctatagtagacttgcaagatccagcagycaaaaaaggaagagagccagagt	966
Db	898	TCAATCCGATTAGCCGACTGGCCCGACAGATCCAGCAGGCAAAAAAGGAGAAGGAGCCAGAGT	957
QY	967	acatgtccttacagaagcaggtcttcacagctgcgagggattgtgacgttaag	1026
Db	958	ACACGCTCCTCACAGAGCGAGGCTCCCGCGCGCAGGAGTTGTGATGACGTTGAAG	1017
QY	1027	ttggscatcacactgcagagaagagtggttaccaataagaaggtgycccaagcgtatgtg	1086
Db	1018	TTGGAACACACACTGCAGAGAAGAACGGGACCAACAAGAAGTTGGCCAAGCGCAATGCAAG	1077
QY	1087	ctgagaacatgctgtagatctctggygttccaagttcccagycgcagcctgcgaagccag	1146
Db	1078	CCGAGAACATGCTGAGAGTCTTGTTTCAAAAGTCCCGCAGCGGCACGCCCAACCCG	1137
QY	1147	cactcaaatcagaagaagaagactccagtaagaanaaccaaggaagcgaagaaagtaacgt	1206
Db	1138	CACCTCAAGTCAGAGAGAGAACACCCTAAAGAAACCAAGGGGATGGAAGAAAAGTAACTT	1197
QY	1207	tttttgaacctagccctgyggtatgnaaatgnaactagtaacaaggaagcagagtccaaga	1266
Db	1198	TTTTGAACCTGGCTGTGGGGATGAAAAATGGACTAGTAATAAAGAGATGAGTTCAGGA	1257
QY	1267	tgccatatcttagccatcacgagctgcccagctgnaattctcccacgtgtgcggaagtgtg	1326
Db	1258	TGCCTTATCTAAGTCAATCAGCAGCTGCTGCTGGAATCTTCCCATGTGTGCCGAGCTCG	1317
QY	1327	cccagagctgtcggygttaagtaaggacacacacaaagaattcaccaagycagctccaa	1386
Db	1318	CCCAGGCTGTAGAGGTAGTCAAGAGACATCACACCAAGAATTTTACCAGGGCAGCTCCGA	1377
QY	1387	atcctgccaagcgaaggttaactgcccattgatatagcccgaagttgtgtacggygcaact	1446
Db	1378	ATCCTGCCAAGGCCACCGTAACTGCCATGATAGCCCGACAGTGTGTATGGGGCACCCT	1437
QY	1447	cgcccaagccgagaccattttaaagatlaacatctcttcaggccaagtaaccocatygac	1506
Db	1438	CGCCCAACAGCCGAGACCAATTTAAAGAATAACATCTCTTCAGGGCCACGTACCCCATGGAC	1497
QY	1507	ctcgcaactagacctctgagcaactgtactaccttcacagagcccaaggaattccaagttg	1566
Db	1498	CTCTCAGGAGACCCTCTGACCAACTGGACTATCTTCCAGAGTCCAGGATTCACAGTTG	1557
QY	1567	aatacaagaatttccaagaacaacaagaagagtglatctcatcaactgtcct	1626
Db	1558	AATACAAAGACTTCCCAAAAACAACAAGAAGCAATTTGTATCTTATCAATGCTCCT	1617
QY	1627	cacagccgcctctcgtcaagtcatgycatcgcgcaaggaatgtgagtcctgttcatgatatgg	1686
Db	1618	CTCAGCCACCTCTGATCAGCCCATGGTATCGGCAAGGATGTGAGTCTTGCCATGATATGG	1677
QY	1687	ctgcactgaacattttaaagctgtctgtctgagttggaaccaacagagcacagagatgycaa	1746

Db	1678	CTGCGCTGAACATCTTAAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAA	1737
QY	1747	gaacaggaatgagccagtttcagcgtgcgaggtgctgaacctttctgycacaac	1806
Db	1738	GAACAGGAACGAGCCACTGTGTGTGTGGAGGTGCTGAACCTTTCTGGCCATGAAC	1797
QY	1807	cattataaaa-cccaacatatatactgaaatactgagaactgcttltgaaatlttggat	1865
Db	1798	CATTATAAATCCCAACATATATACTGAAATATCTGA-AACTGCTTGAATAATTGGAAT	1856
QY	1866	atctgataactccagttggtccaagaca	1892
Db	1857	TTCTGATACCTCCAGTGGGCCGAGAGA	1883

RESULT	5				
LOCUS	AF061938				
DEFINITION	AF061938	3217 bp	mrna	PRI	04-MAR-1999
		Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.			
ACCESSION	AF061938				
VERSION	AF061938.1	GI:4335946			
KEYWORDS	.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 3217)				
AUTHORS	Wickham, L., Duchaine, T., Luo, M., Nabi, I. R. and DesGroselliers, L.				
TITLE	Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum				
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)				
MEDLINE	99147057				
REFERENCE	2 (bases 1 to 3217)				
AUTHORS	DesGroselliers, L., Wickham, L. and Luo, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900				
	Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada				
FEATURES	Location/Qualifiers				
Source	1..3217				

/organism="homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="20q13.1"
 1..3217
 /gene="STAU"
 363..1853
 /gene="STAU"
 /function="RNA binding protein"
 /note="alternatively spliced transcript"
 /codon_start=-1
 /product="stau protein"
 /protein_id="AADI7530.1"
 /db_xref="GI:4335947"
 /translation="MKGKKPMKPYDPSRMOSTYNNMRGAYPPRYFPVPVPL
 LYQVELSVGGQDFNGKGTROAKHDAAKALRILQNEPLPERLEVNGRESEENLNK
 SEISOVFETALKRNLPVNFEVARESPHMKNFVTKVSVGEFGEGSKISKNA
 AIAVLEELKKLPLPAVERVKPRIKKTPIVKPQTSPEYGGINPISRLAQIQAKK
 EKEPEYTLTERGLPRRREFVMQVKGNNHTAEGTGNKKVAKRNAENMLEITGFKVP
 QROPTPALKSEKTPIKKPGDGRKVTFFEEPGSDENGTSNKEDEFMPYLSHOOLPA
 GILPMVPEVAQAQVVSQGHDKTDFTRAAPNPAKATVTAMIARELLYGTSPTAETILK
 NNISGHHVPHGPLTRPSEOLDLTSRVQGFQVEYKDFPKNNKNEFVSLINCSSQPLIS
 HGIGKDYESCHDMAALNLIKLLSELDOOSTEMPRTGNGPMSVCGRC"

Query Match	44.4%;	Score 1269.8;	DB 10;	Length 3217;
Best Local Similarity	81.1%;	Pred. No. 0;		
Matches 1590;	Conservative	0;	Mismatches 277;	Indels 94;
				Gaps 6;

[illegible]

D	1079	GATGCAGGTTGAAGGTTGGAAACCACACTGCAGAGAAGAACGGGCGACCAACAAGAAGTGGC	1138
Q	1073	caagcgtaatgctgctgtgagaacatgctggagatccctygggttcaaagttccccagcgca	1132
D	1139	CAAGCGCAATGCAGCCGAGAACATGCTGGAGATCCTTGTTCAAGTCCCGCAGCGGCA	1198
Q	1133	gcctgccagccagcactcaaatcagagaagaagactccagtlaaagaanaaccagagacgg	1192
D	1199	GCCCACCAACCCCGCACTCAAGTCAGAGGAGAAGACACCCATTAAGAAACCAGGGGATGG	1258
Q	1193	aagaaagtaacglttlttgaacctagccctgggagtgaanaatgnaactaglaacaagga	1252
D	1259	AAGAAAGTAACTTTTGAACCTGGCTCTGGGGATGAAAAATGGGACTAGTAATAAGA	1318
Q	1253	cgaggaqtcaagatgacctatcttagccatcagcagctgcccagctgyaattctcccat	1312
D	1319	GGATGAGTTACGATGCTTATCTAAGTCATCAGCAGCTGCCCTGCTGGAATTCTCCCAT	1378
Q	1313	ggtgcggaagtgcccaagctgtcggggttagtcaagyaacacacaccaaagattcac	1372
D	1379	GGTGCCCGAGGTGCCCCAGGCTGTAGGAGTTAGTCAAGACATCACACCAAGATTTTAC	1438
Q	1373	caaggcagctccaatccctgcgaagyaacgylaactgcatgatagcccagagtggt	1432
D	1439	CAGGCAGCTCCGAATCTCTGCCAAGGCCACGTAACCTGCATGATAGCCCGAGAGTTGT	1498
Q	1433	gtacggggcacctcgcccacagccgagaccatttaaagatlaacatcttcagygca	1492
D	1499	GTATGGGGGCACCTCGCCCCACAGCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCA	1558
Q	1493	cgtaccccatggaacctgcactagaccctctyagcaactgttactaccttccagagccca	1552
D	1559	CGTACCCCATGACCTCTCACGAGACCCCTGTGAGCAACTGGACTATCTTCCAGAGTCCA	1618
Q	1553	gggatccaggttgaatacaaaagatttcccagaacaacaagaagagtgltatctc	1612
D	1619	GGGATTCAGGTTGAATACAAAGACTTCCCAAAAACAACAAGAACGAATTGTATCTCT	1678
Q	1613	catcaactgctctctcacagccgcctctgctcaagtcatgycatcgycgaagatgtgagtc	1672
D	1679	TATCAATTCCTCTCTCAGCCACCTCTGATCAGCCATGGTATCGCAAGCATGTGGAGTC	1738
Q	1673	ctgtcatgatatgctgctgcactgaaacattttaaagctgctgtctgagttygaccaacagag	1732
D	1739	CTGCCATGATATGCTGCGCTGAACATCTTAAAGTTGCTGTCTGAGTTGGACCAACAAG	1798
Q	1733	cacagagatgccaaagaaacaggaatatgaccagttcagcgtgcgggagtgctgaacct	1792
D	1799	TACAGAGATGCCAGAAACAGGAACGACCAATGCTGTGTGTGGAGGTGCTGAACCTT	1858
Q	1793	tctctgcccacaaccattataaaa-cccacatatatactgnaaatactgaaactgct	1851
D	1859	TTCTGGCCATGAACCATTAATAAATCCCAACATATATACTGAAAAATACATGA-AACTGCTT	1917
Q	1852	tgaaaatttgaatatctgataactccagtgtyggccaagaca	1892
D	1918	TGAAATTTGGAATTTCTGATACCTCCAGTGGGCGGAGAGA	1958

RESULT	6		
AF061939			
LOCUS	AF061939	3506 bp	MRNA
DEFINITION	Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced, complete cds.		PRI 04-MAR-1999
ACCESSION	AF061939		
VERSION	AF061939.1	GI:4335948	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 3506)		
AUTHORS	Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.		

TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	source
Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum	Mol. Cell. Biol.	19 (3), 2220-2230 (1999)	2 (bases 1 to 3506)	Desrosiers, L., Wickham, L. and Luo, M.	Direct Submission	Submitted (29-APR-1998)	Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada	Location/Qualifiers
			1. .3506					
			/organism="Homo sapiens"					
			/db_xref="taxon:9606"					
			/chromosome="20"					
			/map="20q13.1"					
			1. .3506					
			/gene="STAU"					
			409. .2142					
			/gene="STAU"					
			/function="RNA binding protein"					
			/note="alternatively spliced transcript"					
			/codon_start=1					
			/product="staufen protein"					
			/protein_id="AAD17531.1"					
			/db_xref="GI:4335949"					
			/translation="MSQVQVQVQNPNSALSSQILNKNSLSQPLMSIPSTSSLPSE ENAGRPIONSALPSASITSTSAAESITPEVELNALCMKLGKTPMYKPYDPSRMOST YNYNMRGAYPPRYFPFPVPPPLLYVELSVGGQFNGKGTROAKHDAKALRIL QNEPLPERLEVNGRESEENLNKSEISOVEIALRNLVNFVARESGPMMKNFVT KVSVEFVGEGEKSKKISKNAIAVLEELKPLPLPAVERVKPRIKKTPIVKPQ TSPEYGGINPISRLAIOIOAKKEKEPEYTLTLTERGLPRREFVMQVKGHNHTEGTS TNKKVAKRNAENMLEILGFKVPORQPIKALKSEKTPIKPGDGKRYTFEPEGSGD ENGTSNKEDFERMPYLSHQQLPAGILPMVPEVAQAVGSQGHKTKDTRAPNPAKAT VTAMIAIRELHYGTSPTAETILKNNISGHVPHGPLTRPSEOLDLTVROGFOVEYKD FPKNKNNEFVSLINCSQPLPLISHGIGKDYVESHDMALNLIKLSLDOOSTEMPR GNGPMSVCGRC"					
BASE COUNT	965 a	875 c	790 g	876 t				
ORIGIN								
Query Match	44.0%;	Score 1256;	DB 10;	Length 3506;				
Best Local Similarity	86.7%;	Pred. No. 0;						
Matches 1420;	Conservative	0;	Mismatches 210;	Indels 8;	Gaps 3;			
QY	262	cagaagatagctctaccattgagctcaatgacatgcatgltgtgaaactgaaagaac	321					
Db	611	CAGAAAGCATTAACCCCTACTAGTAGAATAATGCACTGTGCATGAACCTTGAAAAAAC	670					
QY	322	caatgtataagcccggtgagccctcaactctcgatgacgtccactacagctatgcatgc	381					
Db	671	CAATGTATTAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATACTACAACATGA	730					
QY	382	gtgaggtgcctatccccccagatacttaccacattccagttccaccttactctacc	441					
Db	731	GAGGAGGTGCTTATCCCCCGAGGTACTTTTACCATTTCACAGTTCCACCTTACTTATC	790					
QY	442	aagttgagctctccgtggtggtgacagcagtttaatggaagaagaatgagaccaccg	501					
Db	791	AAGTGAAGCTTCTGTGGAGGACAGCAATTTAATGGCAAGGAAGACAAGACAGGCTG	850					
QY	502	tgaaacacgatgccccctgccccgtgctgaggaactctgcaagatgaaacccctgccagaa	561					
Db	851	CGAAACACGATGCTGCTGCCAAGCGCTTGAGGATCCTGCAGATGAGCCCCCTGCCAGAGA	910					
QY	562	ggttgaggttaaatggaagagaagcagaggaagaagaacctcaataaatcgaaataagcc	621					
Db	911	GGCTGAGGTGAATGGAAGAGATCCGAAGAAGAAATCTCAATAATCTGAATAAAGTC	970					
QY	622	aagtgtttgaattgctgctggaagcggaaatttgctctgtgaatttgaggtggtccggaga	681					
Db	971	AAGTGTGAGATTGCACTTAACGGAACCTTGCCCTGTGATTTTCGAGGTGGCCCCGGAGA	1030					
QY	682	gtggtccaccacacacatggaagaacttgtgaccagggttccagtttggtggaattgttaggg	741					

Db	1031	GTGGCCCAACCCACATGAAGACTTTGTGACCACAGGTTTCGGTTGGGAGTTGTGGGG	1090
QY	742	aaggaagaaagggaaaaagcaagaagatctccaagaagaatgcgycagggctgtctcgaagc	801
Db	1091	AAGGTGAAGGGAAGCAAGAGATTTCAAAGAAAAATGCCGCCATAGCTGTTCTTGAGG	1150
QY	802	agcttaaggagctgcacacccctccctcgtctgtgagcgagtgaagcccaaatcaagaaga	861
Db	1151	AGCTGAAGAAGATTACCGCCCTGCTGCTCAGTTGAACGAGTAAAGCCTAGAAATCAAAAGA	1210
QY	862	aaagtcagcccaacctgca-----agacagccccggatctatgyccaagggatgaactcta	915
Db	1211	AAACAAAAACCCTAGTCAAGCCACACAGACAGCCAGAAATATGGCCAGGGATCAATCCGA	1270
QY	916	ttagtagactctgcacagatccagcagggcaaaaaaggaagaagagccagagtacatgctcc	975
Db	1271	TTAGCCGACTGGCCCGAGATCCAGCAGGCCAAAAAGAGAGAGAGCCAGAGTACCGCTCC	1330
QY	976	ttacagaacgaggtcttcacagtcgcagggagtttgtatgcaggtlaaaggttgagcatc	1035
Db	1331	TCACAGAGCGAGGCTCCCGCCGCGCAGGAGTTGTGATGCAGGTGAAGGTGGAACC	1390
QY	1036	acaatgcagaagagatggtgtaaccaataagaagttgccaagcgtaatgctgtcgaagaca	1095
Db	1391	ACACTGCAGAAGGAACGGGCAACCAACAAGAAGTGGCCAGCGCAATGCAGCCGAGACAA	1450
QY	1096	tgcctgagatcctggtgttcaaaagttccccaagcgacgctgcgaagccaagcatcaaat	1155
Db	1451	TGCTGGAGATCCTTGGTTTCAAGTCCCGCAGCGGCGCACCCCAACCCCGCACTCAAGT	1510
QY	1156	cagaagagaagactccagtaagaaaaacaggaagacggaaggaagtaacgtttttgaaac	1215
Db	1511	CAGAGGAGAAAGACACCCTAATAAGAAACAGGGAGTGAAGAAAGTAACCTTTTGAAC	1570
QY	1216	ctagccctggggatgaaaaatggaactagtaacaaggaagaggttcagagatgcctatc	1275
Db	1571	CTGGCTCTGGGGATGAAAAATGGGACTAGTAATAAAGAGAGATGAGTTCAGGATGCCTTATC	1630
QY	1276	ttagcatcagcagctgcagctggaattctcccaatggtgcggaagttgcccagagctg	1335
Db	1631	TAGTTCATCAGCAGCTGCCCTGCTGGAATTCTTCCCATGTGTCCCGAGGTCCGCCAGGCTG	1690
QY	1336	tcgggggttagtcaaggaacacacacccaagaatttcaccaagggcagctcccaatccctgcca	1395
Db	1691	TAGGAGTTAGTCAAGGACATCACACCAAGATTTTACCAAGGCGAGCTCCGAATCTTGCCA	1750
QY	1396	aggcaacggttaactgccaataagcccgagagttgttgaacgggggcaacctcgcccacag	1455
Db	1751	AGGCCACGGTAAGTGCATGATAGCCCCGAGAGTTGTGTATGGGGGCACCTCGCCACAG	1810
QY	1456	ccgagaccattttaagagtaacatctcttcagyccaagtaaccaatgacactcgcacta	1515
Db	1811	CCGAGACCATTTTAAAGATATAACATCTCTTCAGGCCACGTAACCCCATGAGACCTCTCACA	1870
QY	1516	gacctctgagcaactgttactactttccagagcccagggatctcaggttgaatacaaaag	1575
Db	1871	GACCTCTGAGCAACTGGACTATCTTTCAGAGTCCAGGATTCACAGGTTGAATACAAAG	1930
QY	1576	attttcccaagaacaacaagaacgagtggtatctctcaacaaactgctcctcacagccgc	1635
Db	1931	ACTTCCCAAAAAACAACAAGACGAATTTGTATCTCTTATCAATTGCTCCTCAGCCAC	1990
QY	1636	ctctcgtcagtcatagtcagtcaggcaaggtgtgtgagtcctgtcatgatatagtgtcgaatga	1695
Db	1991	CTCTGATCAGCCATGGTATCGGCAAGGATGTGAGTCTCTGCCATGATATGGCTGCGCTGA	2050
QY	1696	acattttaaagctgtgtctgtagttggaaccaacagagcaacagagatgccaaagaacagaa	1755
Db	2051	ACATCTTAAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAGGAA	2110
QY	1756	atggaaccagtttcagcggtgcgggagagtgctgaacctttctcggccacaacattataaa	1815

Db 2111 ACGGACCAATGTCGTGTGTGGAGGTGCTGAACCTTTTCTGTGCCCATGAACCATATATAA 2170

QY 1816 a-cccacatatatactgaaataactgagaactgcttggaaaatttggaaatatctgataa 1874

Db 2171 ATCCCAACATATATTACTGAAAATACTGA-AACTGCTTTGAAAATTGGAAATTCTGTATAC 2229

QY 1875 ctccagtgggccaagaca 1892

Db 2230 CTCACGTGGCCGAGAGA 2247

RESULT 7

AF061941 3349 bp mRNA PRI 04-MAR-1999

LOCUS AF061941

DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.

ACCESSION AF061941

VERSION AF061941.1 GI:4335952

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3349)

AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.

TITLE Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum

JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)

MEDLINE 99147057

REFERENCE 2 (bases 1 to 3349)

AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.

TITLE Direct Submission

JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada

FEATURES

source

1. 3349

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/map="20q13.1"

1. 3349

/gene="STAU"

495. 1985

/gene="STAU"

/function="RNA binding protein"

/note="alternatively spliced transcript"

/codon_start=1

/product="staufen protein"

/protein_id="AAD17533.1"

/db_xref="GI:4335953"

/translation="MKLGKKPMYKPVDPYSRMOSTYNYNMRGGAYPPRYFYPPVPPL

LYOVELSVGGQOENGKGTQGAHDAAKALRIQNEPLPERLEVNGRESEENLNK

SEISOVEFIALKRNLPVNFVARESGPPHMKNFVTQVSGEFGVEGEGSKISKKNA

AIAYLEELKKLPPLPAVERVPRIKKTKPIVKPQTSPEYGGINPISRLAQIOQAKK

EKEPEYTLTERGLPRRREFVMQVGVNHTAEGTGNKKVAKRNAEENMLEILGFKVP

QROPTKPKLSESEKTPIKKPGDGRKVTFFEPGSGDENGTSNKEDEFMPYLSHQQLPA

GILPMVPEVAQAVGVSOGHHTKDFTRAAPNPAKATVTAMIARELLYGTSPTAETILK

NNISSGHVPHGPLTRPSEQLDLRSVQGFQVEYKDFPKNNKNEFVSLINCSSOPPLIS

HGIGKOVESCHDMAALNIIKLSELDOOSTEMPRTGNGPMSVCGRC"

BASE COUNT 913 a 838 c 777 g 821 t

ORIGIN

Query Match 43.9%; Score 1255; DB 10; Length 3349;

Best Local Similarity 86.7%; Pred. No. 0;

Matches 1419; Conservative 0; Mismatches 210; Indels 8; Gaps 3;

QY 263 agaaagtatagtctaccatctgagctcaatgcaactgtgtgtgaactggaagaaacc 322

Db 455 AGAAAGCATAAACCCCTACTGTAGAACTAATGCACCTGTGCATGAACCTTGAAAAAAC 514

QY 323 aatgtataagcccggtgaccctcactctcgtgagtgcagtcacactacagctatgcatgcg 382

Db 515 AATGTATAAGCCTGTGTGACCCCTTACTCTCGGATGCAGTCCACCCTATTAACATACATGAG 574

QY 383 tggaggtgcctatccccccagatacttccaccattccagtcacccttactctacca 442

Db 575 AGGAGGTCTTATCCCCGAGGFACTTTTACCCATTTCAGAGTTCACCTTTACTTATCA 634

QY 443 agttgagctctccgtggtcgagacagcagtttaatggtgaagaaagatgagaccacct 502

Db 635 AGTGAACCTTCTGTGGGAGGACAGCAATTTAATGGCAAAAGGAAGCAAGACAGCTGC 694

QY 503 gaaacacgatgccccctgccccgtgcgtgagactctgcagagtgaaccccccgcaagaag 562

Db 695 GAAACACGATGCTGCTGCCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCCTGCCAGAGAG 754

QY 563 gttggaagttaatggaagagaagcagaggaagaaacacctcaataatcgyaataagcca 622

Db 755 GCTGGAAGGTGAATGGAAGAAATCCGAAAGAAATACTCAATTAATCTGAATAAGTCA 814

QY 623 agtgttgaatlgcgtgaaagcgaatttcgctgtgaatttgaggtgcccggagag 682

Db 815 AGTGTGAGATTGCACTTAAACGGAACCTGCTGTGAATTTGAGGTGGCCCGGAGAG 874

QY 683 tggccccaccacacatgaagaactttgtgaccagggttcagttgggaaattgtagggga 742

Db 875 TGGCCACCCACACATGAGAACTTTGTGACCAAGCTTTCGTTGGGGAGTTGTGGGGA 934

QY 743 aggaagaaggaaagaagaagatctccaagaagaatgcgcccagggtgtctgagca 802

Db 935 AGGTGAAGGGAAGAAAGCAGAGAAGATTTCAAAGAAATAATGCCCATAGCTTCTTGAGGA 994

QY 803 gcttaggaagctgccaccctccctcgtgtgagcgaagtgaagcccgaatcaagaaga 862

Db 995 GCTGAAGAAGTTACCGCCCTGCTGTCAGTTGAAGTAAGGCTTGAATCAAAAAAGAA 1054

QY 863 aagtcagccaccctgca-----agacagcccggattatggtccaaaggatgaatcctat 916

Db 1055 AACAAAACCATAGTCAAGCCACAGACAGACCAGAAATATGGCCAGGGATCAATCCGAT 1114

QY 917 taqtagacttgacagatccagcagcaggaagaaaggaagagcagagtatcatgtcct 976

Db 1115 TAGCCGACTGGCCCGACAGTCCAGCAGGCAAAAAAGGAAGAGCCAGACTACCGCTCCT 1174

QY 977 tacagaacgaggtcttcacacgtcgcagggagttgtgatgcaggtaaaggttggcatca 1036

Db 1175 CACAGAGCGAGGCGCTCCCGCGCCGACGAGGAGTTGTGATGCAGGTGAAGTTGGAACCA 1234

QY 1037 cactgcagaagagagtggttaccaataagaagtggtccaaagcgtaatgtctgtagaacat 1096

Db 1235 CACTGCAGAAAGAACGGGCAACCAAGAGGTGGCCCAAGCGCAATGCAACCGAGAACAT 1294

QY 1097 gctggaagatcctggtgtcaaaagttccccagggcagcctgccaaagccaacccaatc 1156

Db 1295 GCTGGAGATCCTTGGTTTCAAGTCCCGCAGCGGCACGCCAACCAACCCGCACCTCAAGTC 1354

QY 1157 agaagaagaagactccagtaaaagaaccaggaagcgaagaaagaaagtaacgtttttgaacc 1216

Db 1355 AGAGGAGAAGACACCCATTAAGAAACACGAGGGAATGAAGAAAGTAACCTTTTGAACC 1414

QY 1217 tagccctggggtatgaaaatggaactagtaacaaggaagagagttcagagatgcctatct 1276

Db 1415 TGGCTCTGGGATGAATAATGGGACTAGTAATAAAGAGATGAGTTTCAGAGATGCTTATCT 1474

QY 1277 tagccatcagcagctgcagcagctggaattctcccatggtgccgaagttgccaggtgt 1336

Db 1475 AAGTCATCAGACAGCTGCCCTGCTGGAATTTCTCCCATGCTGCCGAGGTGCGCCAGGCTGT 1534

QY 1337 cggggttagtcaaggaacacacaccaaagatttccaccagggcagctccaatctctgcaa 1396

Db 1535 AGGAGTTAGTCAAGGACATCACACCAAGATTTTACCAAGGGCAGCTCCGAATCTTGCCAA 1594

QY 1397 ggcaacggttaactgcatgatagcccgagaggtgtgttacggygggcaacctgcgccacagc 1456

Db 1595 GGCCACGGTAACCTGCATGATAGCCCGAGAGTTGTGTATGGGGGCACTCGCCACACAGC 1654

QY	1457	cgagaccattttaaagaaglaacatctctcagggccagctaacccatggaacctgcgactag	1516
Db	1655	CGAGACCATTTTAAAGAATAACATCTCTTCAGGGCCACGTAACCCATGGACCTCTCAGCAG	1714
QY	1517	accctctgagcaactgtactaccttccagagcccaaggatccaggltgaatacaaga	1576
Db	1715	ACCCTCTGAGCAACTGGACTATCTTTCAGAGTCCAGGAGATCCAGGTGAATACAAAGA	1774
QY	1577	tttcccaagaacaacaagaacgagtgtatctctcatcaactgctctcacagccgc	1636
Db	1775	CTTCCCCAAAAACAACAAGAAGAAATTTGTATCTCTTATCAATGTCTCCTCAGCCACC	1834
QY	1637	tctcgtcagtcacgycalcgcgcaagatgtgagtgctctgtcatgatatgctgcactga	1696
Db	1835	TCTGATCAGCCATGTGTATTCGGCAAGATGTGGAGTCTCCATGATATGGCTGGCTGAA	1894
QY	1697	cattttaagctgctgtctgtagttggaaccaacagagcacagagatgccaagaacagaaa	1756
Db	1895	CATCTTAAGTGTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAGGAAA	1954
QY	1757	tggaccagtttccagcgttcggygagtgctgtyaaccttctctgcccacaaaccattataaa	1816
Db	1955	CGGACCAATGTCTGTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAACCATTTATAAA	2014
QY	1817	-cccacatatatacttgaaaataactgagaaactgcttggaaaatttggaaatatctgataac	1875
Db	2015	TCCCAACATATATTACTGAAATACTGA-AACTGCTTTGAAAATTGGAAATTCTGATACC	2073
QY	1876	tccagtgggccaagaca 1892	
Db	2074	TCCAGTGGGCCGAGAGA 2090	

RESULT	8	
AK025519		
LOCUS		
DEFINITION	AK025519	3194 bp mRNA
ACCESSION	AK025519	PRI
VERSION	AK025519.1	GI:10438059
KEYWORDS	oligo capping; fls (full insert sequence).	
SOURCE	Homo sapiens hepatoma cell_line:HepG2	CDNA to mRNA, clone_1lb:HEP clone:HEP02379.
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (sites)	
TITLE	Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.	
JOURNAL	NEDO human cDNA sequencing project	
REFERENCE	Unpublished (2000)	
AUTHORS	2 (bases 1 to 3194)	
TITLE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.	
JOURNAL	Direct Submission	
TITLE	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416).	
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	
FEATURES	Location/Qualifiers	
source	1..3194	
	/organism="Homo sapiens"	

```
misc_feature
1..3194
/note="highly similar to AF061938 Homo sapiens staufer
protein (STAU) mRNA"
BASE COUNT      894 a      787 c      732 g      781 t
ORIGIN
```

[illegible]

QY 812 gctgccaccctccctgctgtgtgagcgagtgtaagcccagaatccaagaagaaagtcagcc 871
Db 841 GTTACCGCCCTGCTGCGAGTTGAACGAGTTAAAGCCTAGAAATCAAAAAAGAAACAAACC 900
QY 872 -----cacctgcaagacagccccgaltatgccaaggatgaatcctatagtagact 926
Db 901 ATAGTCAAGCCACAGACAAGCCAGAAATATGGCCAGGGATCAATCCGATTAGCCGACTG 960
QY 927 gcacagatccagcagcgaaaaaagagaagagccagagtlacatgctccttacagaacga 986
Db 961 GCCCAGATCCAGCAGGCAAAAAAGAGAGAGCCAGAGTACACGCTCTCCACAGAGCGA 1020
QY 987 ggtctccacgtcgcaaggagtttgtatgcaggttaaagttggcatcacactgcagaa 1046
Db 1021 GGCTTCCCGCCGCGAGGAGTTGTGATGCGAGTGAAGTTGGAACCAACACATGCGAGAA 1080
QY 1047 ggaagtggtaaccaataagaaggtggccaagcgtaatgctgtgagaacatgctggagatc 1106
Db 1081 GGAACGGGCGACCAACAAGAAGGTGGCCAGCGCAATGCAGCCGAGAACATGCTGAGATC 1140
QY 1107 ctggggttcaaagttcccccaggcgagcctgccaagccagcactcaatcagaagaag 1166
Db 1141 CTTGCTTCAAAAGTCCCGCAGCGCGACGCCACCAAAACCCGACTCAAGTCAGAGAGAAG 1200
QY 1167 actccagtaaagaaccagagagcagcgaagtaacglttttgaaacctagccctggg 1226
Db 1201 ACACCCATAAAGAAACCGAG -GATGGAAGAAGTAACCTTTTGAACCTGGCTCTGGG 1259
QY 1227 gatgaanaatggaactagtaacaagcagcagaggttcagatgcctatccttagccatcag 1286
Db 1260 GATGAANAATGGGACTAGTAATAAGAAGGATGAGTTCAAGATGCCCTTATCTAAGTCATCAG 1319
QY 1287 cagctgcagctggaattctcccatggtgctcggaagagttgcccaagcgtcggggttagt 1346
Db 1320 CAGCTGCCCTGCTGGAATTTCTCCCATGGTGCCCCAGAGGTCGCCAGGCTGTAGAGTTAGT 1379
QY 1347 caagcagaccacaccaaagattccaccagcagctccaatcctgcacaaagcgaacgyla 1406
Db 1380 CAAGGACATCACACCAAGAATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGGTA 1439
QY 1407 actgccatgatagcccagaggttgtgtacaggggcacctgcgccacagccgagaccatt 1466
Db 1440 ACTGCCATGATAGCCCCGAGAGTTGTTGATGGGGCACCTCGCCACAGCCGAGACCATT 1499
QY 1467 ttaaagagtaacatctcttcaggccacgtaccatgacctgcactagaccctctgag 1526
Db 1500 TTAAGAATAAACACCTCTTCAGGCCACGTAACCCCATGGAACCTCTCACAGAGACCCCTGAG 1559
QY 1527 caactgtactactcttcacagagcccaaggatccaggttgtaatacaaaagatttcccaag 1586
Db 1560 CAACTGGACTATCTTCCACAGATCCAGGATTCAGGTTGAATACAAAGACCTCCCCAAA 1619
QY 1587 aaccaacaagaacgagtggtatctctcatcaactgctcctcacagccgctcctcgtcagt 1646
Db 1620 AACCAACAAGAAGAAATTTGATCTCTTATCAATGCTCTCTCAGGCCACCTCTGATCAGC 1679
QY 1647 catggcaltcgcaagagtggtgagtcctgtcatgatatggtcgtcactgaacattttaaag 1706
Db 1680 CATGGTATCGGCAAGATGGAGTCCCTGCCATGATATGGCTGGCTGGAACATCTTAAAG 1739
QY 1707 ctgctgtctgagttgaccaacagagcacagagatgccaagaagaaagaaatggaccagtt 1766
Db 1740 TTGCTGTCTGAGTTGACCAACAAGTACAGAGATGCCAAGAAGAAACGAGACCAATG 1799
QY 1767 tcaagcgtcgaggagtgctgaaaccttttctgccaacaaacattataaaa-cccaacata 1825
Db 1800 TCTGTGTGTGGAGGTGCTGAACCTTTTCTGCGCATGAACCATTTATAAATCCCAACATA 1859
QY 1826 tatactgaaaaatactagagaactgcttgaaaatttggaaatatctgataaactccagtgagc 1885
Db 1860 TATACTGAAAAATACTGA-AACTGCTTTGAAAAATTTGGAATTTCTGATACCTCCAGTGGGC 1918
QY 1886 caagaca 1892

Db 1919 CGAGAGA 1925
RESULT 9
HSA132258
LOCUS HSA132258 3066 bp mRNA PRI 14-MAY-1999
DEFINITION Homo sapiens mRNA for staufen protein, partial.
ACCESSION AJ132258
VERSION AJ132258.1 GI:4572587
KEYWORDS staufen gene; staufen protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3066)
AUTHORS Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
TITLE Interaction of Influenza virus NS1 protein and the human homologue
of Staufen in vivo and in vitro
JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE 99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS Ortin,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,
Centro Nacional de Biotecnología, Cantoblanco, 28049 Madrid, SPAIN
FEATURES
source 1..3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 36..1526
/gene="staufen"
36..1526
/gene="staufen"
CDS /function="dsRNA binding protein"
/codon_start=1
/product="staufen protein"
/protein_id="CAB4082.1"
/db_xref="GI:4572588"
/translation="MKLGKKPMKPYDPYSRMQSTYNMNGAYPRRYFPYFPVPL
LYQVELSYGOQFNGKGKTRQAAKHDAALRLILQNEPLERLEVNGRESEENLNK
SEISQVEFETALRNLVPNFEVARESGPPHMKNFVTKSVGEFVEGEGSKSKIKNA
AIAVLEELKKLPPLPAVERVKPRIKKTKPIVKPOTSPEGOGINPISRLAQIOAKR
EKEPEYTLTERGLPRRREFVMQYVGNHTAEGTGTNKKYAKRNAENMLEILGKVP
ORSHQRTQVGEDTHKETRGWKSNIPLNALGMENGTSMKDEFRMPYLSHOOLPA
GILPMVPEVAQAVGSQGHHTKDTTPAAPNPAKATVTAMARELLYGTSPYAEITLK
NNISSGHVPHGPLTRPSEQLDYSRVOGEQVEYKDFPKNNKNEFVSLINCSSQPLIS
HGIGKDVESHDMALNLIKLSLDQOSTEMPRTGNGPMVSCGRC"
BASE COUNT 891 a 693 c 699 g 783 t
ORIGIN
Query Match 42.3%; Score 1209.2; DB 68; Length 3066;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 1409; Conservative 0; Mismatches 213; Indels 12; Gaps 6;
QY 268 gtaagcttataccattgagctcaatgacactgtgtgaaactggaagaaacccaatgt 327
Db 1 GCATTAACCCCTACTGTAGACTAAATGCACACTGTGATGAAGAACTTGAAAAAACCAATGT 60
QY 328 ataagccgtggaacctcactctcgatgcagtcacactcacagctatgcatgctgag 387
Db 61 ATAAGCCTGTTGACCCCTTACTCTCGGATGCAGTCCACCTATACTACAACATGAGAGAG 120
QY 388 gtgcctatccccccagatacctttacccaattccagttccacaccttactctaccaagttg 447
Db 121 GTGCTTATCCCCCGAGGTACTTTTACCCATTTCAGATTCCACACTTACTTTATCAAGTGG 180
QY 448 agctctccgttgggagcagcagtttaatggaagaaagaaagatgagaccaccgctgaac 507
Db 181 AACTTCTGTGGAGAGACAGCAATTTAATGGCAAGAAAGACAAGACAGGCTGCCAAGAC 240
QY 508 acgatgccctcgcccggtgcgtgagactctgcagagtgaaacccctgccagaagagttgg 567

```
Db 241 ACGATGCTGCTGCCAAAGCGTTGAGGATCTGCAGAAATGAGCCCTCGCCAGAGAGCTGG 300
QY 568 aggtaaatggaagaagcagaggaagaacctaataatcgaaataagccaagt 627
Db 301 AGGTGAATGGAAGAGATCCGAGAGAGAAATCTCAATAATCTGAATAAGTCAAGTGT 360
QY 628 ttgaattgcgcgtgaagcggaaatttgccgtgtgaatttgaggtgccccggagagtgcc 687
Db 361 TTGAGATTGCACCTTAACCGGAACCTTGCTGTGAATTTGAGAGGTGCGCCGAGAGTGCC 420
QY 688 caccacacatgaagaactltgtgaccaggttccagtltgaggaaattgtlaggggaagag 747
Db 421 CACCCACATGAAGAACTTTGTGACCAAGTTTCGTTGGGAGTTGTGGGGAAGGTG 480
QY 748 aaggaaagaagaagaatctccaagaagaatgcgagcgaggtcttctgagcagctta 807
Db 481 AAGGAAAGCAAGAGATTTCAAAGAAATGCGCCATAGCTGTCTTGAGAGAGCTGA 540
QY 808 ggaagctgcacacccctccctgctgtgagcgagtgaaagcccaagaatcaagaagaatgc 867
Db 541 AGAAGTTACCGCCCTGCTGCTGACAGTGAACGAGTAAGCCTAGAAATCAAAAGAAACAA 600
QY 868 agcccaactgca-----agacagcccccagaltatgcccagaggaatgaatcctatagta 921
Db 601 AACCCATAGTCAAGCCACAGACAAAGCCAGAAATATGGCCAGGGGATCAATCCGATTAGCC 660
QY 922 gacttcacagatccagcaggaagaaaggaagagcagagatcatatgctccttacag 981
Db 661 GACTGGCCAGATCCAGCAGGCAAAAAAGAGAAAGAGCCAGAGTACACGCTCCTCACAG 720
QY 982 aacgagctctccagctgcagaggtgtgtatgacgttaaaggttgagcatcacatc 1041
Db 721 AGCAGAGCCTCCCGCCGCGCAGGAGATTGTGATGCAGGTGAAGTTGGAACACACACTG 780
QY 1042 cagaagagtggttaccaataagaagtgcgaagcgtaatgtctgtgaacatgtcgtg 1101
Db 781 CAGAAGGAACGGGACCAACAAGAGAGGTGGCCCAAGCGCAATGCAGCCGAGAATCATGCTG 840
QY 1102 agatcctggtgttcaaaagttcccaagcgagccttgccaagccagcactcaaatcagaag 1161
Db 841 AGATCCTTGTTCAAAAGTCCCGCAGCGCAGCCA--CCAACCCGCACTCAAGTCAGAGG 898
QY 1162 agaagactccagtaagaagaaacagagagcgaaggaagtaa-cgttttltgaacctagc 1220
Db 899 AGAAGACACCCATTAAGAAACAGGCGATGGAAGAAAGTAACCTTTTGAACCTGGC 958
QY 1221 cctgggat-gaaaatlygaactagtaacaagagcagagatcagatgcctatcttag 1279
Db 959 TCTGGGATGGAATAATGGGACTAGTAATAAGAGATGAGTTCAAGATGCCCTTATCTAAG 1018
QY 1280 ccatacagcagctgcagctggaattctccocatggtgccggaagtgtgccaggtgtcg 1339
Db 1019 TCATCAGCAGCTGCTGCTGGAATTCCTCCCATGTGTGCCGAGTGCGCCAGGCTGTAGG 1078
QY 1340 ggttagtcaagagacacacaccaaagaatttcacacagggcagcttccaatctcgtccaagc 1399
Db 1079 AGTTAGTCAAGGACATCACACCAAGATTTTACCAGGGCAGCTCCGAATCTCTGCCAAGGC 1138
QY 1400 aacggttaactgcatgatagcccgagaggtgtgtacggygggacacctcgccacagccga 1459
Db 1139 CACGGTAACCTGCATGATAGCCCCGAGAGTTGTGTATGGGGGCACTCGCCACAGCCGA 1198
QY 1460 gaccatttaagagtaacatctcttcagggccaagtlaccccatgtaacctgcactagacc 1519
Db 1199 GACCATTTTAAAGATAACATCTCTTCAGGGCCACGTACCCCATGACCTCTCACGAGACC 1258
QY 1520 ctctgagcaactgactactcttccagagcccaaggatccaggttgaataacaagaattc 1579
Db 1259 CTCTGAGCAACTGCACTATCTTTCAGAGTCCAGGGATTCAGAGTTGAATACAAAGACTT 1318
QY 1580 tcccaagaacaagaagaagagtggtgtatctctcatcaactgtcctcacaagccgctct 1639
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
Db 1319 CCCCCAAAACAAGACGAATTTGATCTCTTATCAATTCCTCTCTCAGCCACTCT 1378
QY 1640 cgtcagtcatgcatcgcaagagatgtgagtcctgtcatgatatgctgcactgaacat 1699
Db 1379 GATCAGCCATGATTCGGCAAGAGATGTGAGTCTCCCATGATATGTGCTGCGCTGAACAT 1438
QY 1700 ttaaagctgtctgtgagtltgaccacaagcagacagatgccaagaagaagaatg 1759
Db 1439 CTTAAAGTTGCTGTGAGTTGGACCAACAAGTACAGAGATGCCAAGAAACGGAACGG 1498
QY 1760 accagtltcagcgtgcgggaggtgtgtgaaccttctcggccacaacatataaaa-cc 1818
Db 1499 ACCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGGCCATGAACCATTAATAAATCC 1558
QY 1819 caacatatatactgaaaatactgagaactcttlyaaaatttgaatatctgataactcc 1878
Db 1559 CAACATATATACTGAATAATACTGA-AACTGCTTTGAAAATTGGAATTTCTGATACCTCC 1617
QY 1879 agtggccaagaca 1892
Db 1618 AGTGGCCGAGAGA 1631
```

```
RESULT 10
AC063967/c
LOCUS
DEFINITION
Mus musculus clone RP23-41G23, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
AC063967
VERSION
AC063967.3 GI:9972323
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
house mouse.
ORGANISM
Mus musculus
```

```
REFERENCE
AUTHORS
McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Shekher,M., Spiegel,L.A., Toth,K. and Vill,M.D.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 225016)
McCombie,W.R.
Direct Submission
Submitted (22-APR-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Sep 6, 2000 this sequence version replaced gi:8439872.
```

```
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Center project name: RP23-41G23
Center clone name: RP23-41G23
-----
```

```
NOTE: This is a 'working draft' sequence. It currently
consists of 13 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
```

```
1 121343: contig of 121343 bp in length
* 121344 121985: gap of unknown length
* 121986 139930: contig of 17945 bp in length
* 139931 140572: gap of unknown length
* 140573 155472: contig of 14900 bp in length
* 155473 156114: gap of unknown length
```


/protein_id="CAA09037.1"
/db_xref="GI:4138428"
/translation="YFYPFVPPPLYQVELSVGGQDFNGKGMPPVKHDATA"
BASE COUNT 109 a 70 c 105 g 76 t
ORIGIN

Query Match 11.4%; Score 325.4; DB 74; Length 360;
Best Local Similarity 94.2%; Pred. No. 4.5e-77;
Matches 338; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 405 tactttaccattccacccacttactctaccaagtgtgagctctcgtggcgga 464
Db 2 TACTTTACCCATTCCAGTTCACCGCTTACTCTACCAAGTTGAGCTCTCCGTGGAGGA 61
OY 465 cagcagtttaatgggaaagaagatgagacccacctgaacacagatgccccctgcggt 524
Db 62 CAGCAGTTTAATGGGAAGGAAGATGAGACCACTGTGAACATGATGCCACTGCCCGA 121
OY 525 gcgcctgagactctgcagagtgaaacccctgcagaaagtttgaggttaaatggaagaa 584
Db 122 GCCCTGAGGACTCTGCAGAGTGAGCCCCCTTACCAGAGGCTAGAGGTAATGGAAGAGAA 181
OY 585 gcagaagaaagaacctcaataaatgcgaaataagccaagtgttgaaattgcgctgaag 644
Db 182 TCAGAAGAAGAAACCTTAATTAATCAGAAATAAGCCAAGTGTGAGATTGCGCTGAAG 241
OY 645 cgaatttgcttgaaatttgaggtgcccggagagtgcccacacacatgaagaac 704
Db 242 CGGAATTTGCGCTGCAATTTTGAGGTGCGCCGGAGAGTGGCCGCCACACATGAAGAAC 301
OY 705 ttgtgaccaggtttcagttggtggaattgtaggggaaggaaggaagaa 763
Db 302 TTTGTGACCAGGGTTTCAGTTGGGAGTTTGTAGGGGAAGGAGGAAGCAAGAA 360

RESULT 12
AL133174/c 113836 bp DNA 14-JUN-2000
LOCUS Human DNA sequence from clone RP3-470L14 on chromosome 20, complete
DEFINITION sequence.
ACCESSION AL133174
VERSION AL133174.15 GI:8573761
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 113836)
AUTHORS Ramsay, H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquyes@sanger.ac.uk
COMMENT On Jun 20, 2000 this sequence version replaced gi:8388434.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human

chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP3-470L14 is from the library RPI-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pcypac2
IMPORTANT: This sequence is not the entire insert of clone
RP3-470L14. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP3-470L14 is at 113836 in this
sequence. The true right end of clone RPI-15566 is at 100 in this
sequence.

FEATURES
Source Location/Qualifiers
1..113836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP3-470L14"
/clone_lib="RPI-3"
/clone="RPI-3"
3061..3077
/note="Weak data"
misc_feature
9583..95627
/note="Single clone region. pUC subcloned to give 12x
coverage. Assembly confirmed by restriction digest data."
BASE COUNT 33974 a 25115 c 24233 g 30514 t
ORIGIN

Query Match 9.9%; Score 283.6; DB 34; Length 113836;
Best Local Similarity 92.5%; Pred. No. 1.7e-65;
Matches 298; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 1243 gtaacaagacgagaggttcagatgccttatcttagccatcagcagctgccagctgga 1302
Db 38168 GTAATAAGAGAGATGAGTTCAGAGATGCCCTTATCTAAGTCATCAGCAGCTGCCCTGGA 38109
OY 1303 ttctcccatgtgcccgaagttgcccaggtgtcggtttagtcaaggacaccaccca 1362
Db 38108 TTCTTCCCATGTGTCGCCGAGGTGCGCCAGGCTGTAGAGTTAGTCAAGACATCACACCA 38049
OY 1363 aagatttcaccagggcagctcccaatctctgccaaggcaacggttaactgcatgatagccc 1422
Db 38048 AAGATTTTACCAGGGCAGCTCCGAATCTTGCCCAAGGCCACGCTAATGATAGCCCC 37989
OY 1423 gagagttgttgacgggggacacctgcgccacagccgagacatttaagtagaacctc 1482
Db 37988 GAGAGTTGTTGATGGGGGACACCTCGCCACAGCCGAGACCAATTTAAAGATAACATCT 37929
OY 1483 ctccagccacgtaccatcagctgcagctgacactagaccctctgagcaactgtactactt 1542
Db 37928 CTTACGGCCACGTAACCCCATGAGACCTCTGACGAGACCCCTGACCAACTGACTATCTTT 37869
OY 1543 ccagagcccgaggttccaggt 1564
Db 37868 CCAGAGTCCAGGCAATCCAGGT 37847

RESULT 13
AK002152 2968 bp mRNA 22-FEB-2000
LOCUS Homo sapiens CDNA FLJ11290 f1s, clone PLACE1009622, weakly similar
DEFINITION to MATERNAL EFFECT PROTEIN STAUEN.
ACCESSION AK002152
VERSION AK002152.1 GI:7023857
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
clone:PLACE1009622.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,

[illegible]

QY	502	tgaacacgagtgcctcgtgccgtgctgagtgacttcgcagagtgaacccctgcaga	561
Db	754	CTAGACACAATGCTGCATGAAGCCCTCCAAGCACTGCAGAATGAACCTATTCCAGAAA	813
QY	562	ggltgagtgtaaatggaagagaaca-----gaggaagaaaacctcaataat	609
Db	814	GATCTCCTCAGAATGTGTAATCAGAAAGGATATGATGATGACAAAGATGCAAATAGT	873
QY	610	cggaataagccaagtgttgaatttgcgtgaagcggaatttgccttgtaatlttagg	669
Db	874	CTGAGATCACGCTTAGTGTTGAATAATGTCTCTGAAGCGAATAATGCCCTGTACAGTTTGAGG	933
QY	670	tggcccggagagttgccccaccacacatgaagaactttgtgaccagggtttcagttggg	729
Db	934	TATTTAAGAAAGTGGACCACCATATATGAATAAGCTTTGTTACTCGAGTGTACAGTAGCAG	993
QY	730	aatttgtaggagaagaaaggaagaaagcaagaagatctccaagaagaatgcygcccagg	789
Db	994	AGTTCTCTGCAGAGGAGGAAGAAATAGCAAAAAAAGCTCCCAAGAAAGCGCGTCGACCA	1053
QY	790	ctgttctgagcagctagaaggctgcacccctccctctgtlygagcgagtgaagccca	849
Db	1054	CCGTCCTTACAGAGCTTAATAAACTTCCACCTCTTCTCTGTGTGGAAAAAGCCAAACATAT	1113
QY	850	gaatcaagaagaanaagtcagcccacctgcagaacagcccgattatlgccaagsgatga	909
Db	1114	TTTTTAAAAAACGCCCTTAACAATAGTAAAGGCCGACCAAGATATATGCCCCAAGGATGA	1173
QY	910	atcctatagtagacttgcacagatccagcagcgcaaaaaagagaagagccagagtac	969
Db	1174	ACCCATTAGCCGCCCTGGCCAAATTCAACAGAGGCCAAAAAGAAAGAGCCGATTATG	1233
QY	970	tgtccttacagaagaggtcttccacgtgcgagygagttgtgatgcagtgtaaagttg	1029
Db	1234	TTTTGCTTTCAGAAAGAGAAATGCCCTCGACGTGAGAAATTTGTGATGCAGGTGAAGGTAG	1293
QY	1030	ggcatcacactgcagaagaggttggttaccaataagaagtyggccaagcgtatgtctg	1089
Db	1294	GCAATGAAGTTGCTTACAGNAACAGACCTAATATAAAGATAGCCAAAAAATGTGCAG	1353
QY	1090	agaacatgctgagatcctgyggttcaaagttccc	1124
Db	1354	AAGCAATGCTGTTACAACCTGTTATAAAGCATCC	1388
RESULT	14		
LOCUS	HSA9062	4058 bp mRNA	PRI 14-JAN-2000
DEFINITION	Homo sapiens mRNA for 39k3 protein.		
ACCESSION	Y19062		
VERSION	Y19062.1 GI:6634456		
KEYWORDS	39K3 protein.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 4058)		
JOURNAL	Buchner,G., Bassi,M.T., Andolfi,G., Ballabio,A. and Franco,B.		
MEDLINE	Identification of a novel homolog of the drosophila staufer protein		
REFERENCE	in the chromosome 8q13-q21.1 region		
AUTHORS	Genomics 62 (1), 113-118 (1999)		
TITLE	2 (bases 1 to 4058)		
JOURNAL	Franco,B.		
FEATURES	Direct Submission		
source	Submitted (07-JUN-1999) B. Franco, TIGEM, Via Olgettina 58, 20132 Milan, ITALY		
	Location/Qualifiers		
	1..4058		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="teratocarcinoma-neuron"		


```

repeat_region 695. .3450
/note="L1P3 repeat: matches 232. .2841 of consensus"
repeat_region 3451. .6427
/note="L1P2 repeat: matches 3148. .6144 of consensus"
repeat_region 6432. .7143
/note="L1MD1 repeat: matches 5283. .6016 of consensus"
repeat_region 7164. .7316
/note="AluJb repeat: matches 137. .290 of consensus"
repeat_region 7339. .7646
/note="L1MD2 repeat: matches 6016. .6330 of consensus"
repeat_region 7650. .7754
/note="L1MC5 repeat: matches 7608. .7714 of consensus"
repeat_region 7767. .8072
/note="AluY repeat: matches 1. .305 of consensus"
repeat_region 8095. .8136
/note="21 copies 2 mer ca 90% conserved"
repeat_region 8152. .8245
/note="L1MC5 repeat: matches 7478. .7572 of consensus"
repeat_region 8273. .8494
/note="AluJo repeat: matches 84. .305 of consensus"
repeat_region 8503. .8633
/note="L1MC5 repeat: matches 7305. .7446 of consensus"
repeat_region 8712. .8879
/note="84 copies 2 mer tt 72% conserved"
repeat_region 8899. .9190
/note="AluJo repeat: matches 8. .300 of consensus"
misc_feature complement(9869. .10121)
/note="match: STS: Em:L41960; match: STS: Em:L41960"
repeat_region 12531. .13017
/note="L2 repeat: matches 2083. .2505 of consensus"
repeat_region 13018. .13320
/note="L1MA1 repeat: matches 6004. .6303 of consensus"
repeat_region 13321. .14004
/note="L2 repeat: matches 1197. .2083 of consensus"
misc_feature complement(16833. .17221)
/note="match: GSS: Em:AQ558019"
misc_feature 16840. .17277
/note="match: GSS: Em:B65742"
misc_feature complement(16845. .17241)
/note="match: GSS: Em:AQ149727"
misc_feature complement(16848. .17220)
/note="match: GSS: Em:AQ240745 Em:AQ474471"
repeat_region 16886. .17629
/note="MER4D repeat: matches 234. .973 of consensus"
repeat_region 18183. .18407
/note="L1P3 repeat: matches 5928. .6148 of consensus"
repeat_region 18455. .19002
/note="MER41A repeat: matches 1. .554 of consensus"
repeat_region 19289. .19465
/note="MIR repeat: matches 15. .198 of consensus"
misc_feature 19583. .20000
/note="match: GSS: Em:AQ017861"
repeat_region 19875. .20274
/note="MER57B repeat: matches 1. .403 of consensus"
repeat_region 20275. .21084
/note="MER57 internal repeat: matches 1. .806 of consensus"
repeat_region 21078. .22422
/note="MER57 internal repeat: matches 440. .2799 of consensus"
repeat_region 22425. .22764
/note="MER57 internal repeat: matches 3892. .4230 of consensus"
repeat_region 22765. .23041
/note="AluX repeat: matches 8. .286 of consensus"
repeat_region 23042. .24919
/note="MER57 internal repeat: matches 4230. .6114 of consensus"
repeat_region 24939. .25118
/note="L1PBI repeat: matches 5980. .6153 of consensus"
repeat_region 25119. .25422
/note="AluY repeat: matches 1. .304 of consensus"
repeat_region 25423. .25607
/note="L1PBI repeat: matches 5777. .5980 of consensus"

```

```

repeat_region 25635. .25672
/note="19 copies 2 mer ta 78% conserved"
repeat_region 25681. .26070
/note="L1P1 repeat: matches 5397. .5790 of consensus"
repeat_region 26082. .27136
/note="MER57-internal repeat: matches 6489. .7537 of consensus"
repeat_region 27137. .27249
/note="MER57B repeat: matches 1. .114 of consensus"
repeat_region 27250. .28097
/note="L1MC3 repeat: matches 6763. .7631 of consensus"
repeat_region 28123. .28390
/note="L1MC3 repeat: matches 28. .313 of consensus"
repeat_region 28404. .28952
/note="L1MC3 repeat: matches 5376. .5993 of consensus"
repeat_region 28946. .29622
/note="L1M4 repeat: matches 4880. .5566 of consensus"
repeat_region 30851. .31163
/note="AluSc repeat: matches 1. .301 of consensus"
repeat_region 32518. .34668
/note="L1PA14 repeat: matches 3983. .6149 of consensus"
repeat_region 34672. .35189
/note="L1M2B repeat: matches 1. .440 of consensus"
repeat_region 35190. .36499
/note="L1PA14 repeat: matches 2718. .3988 of consensus"
repeat_region 36500. .36876
/note="THE1C repeat: matches 1. .369 of consensus"
repeat_region 36877. .38213
/note="L1PA14 repeat: matches 1264. .2718 of consensus"
repeat_region 38221. .39169
/note="L1PA14 repeat: matches 288. .1240 of consensus"
repeat_region 39455. .39517
/note="L2 repeat: matches 2349. .2404 of consensus"
repeat_region 40251. .40337
/note="MIR repeat: matches 59. .146 of consensus"
repeat_region 41354. .41648
/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region 41650. .41743
/note="47 copies 2 mer aa 62% conserved"
repeat_region 41744. .42318
/note="L1 repeat: matches 4553. .5128 of consensus"
repeat_region 42337. .42391
/note="Alu repeat: matches 251. .302 of consensus"
repeat_region 42360. .42395
/note="Alu repeat: matches 79. .114 of consensus"
repeat_region 42421. .43734
/note="L1 repeat: matches 4019. .5334 of consensus"
repeat_region 43735. .43834
/note="FLAM_C repeat: matches 32. .131 of consensus"
repeat_region 43835. .44621
/note="L1 repeat: matches 3232. .4019 of consensus"
repeat_region 44622. .44802
/note="L1M1 repeat: matches 6121. .6304 of consensus"
repeat_region 44803. .45815
/note="L1 repeat: matches 2126. .3232 of consensus"
repeat_region 45514. .46762
/note="L1M1 repeat: matches 1155. .2794 of consensus"
repeat_region 46760. .47327
/note="L1M1 repeat: matches 59. .642 of consensus"
repeat_region 47325. .48451
/note="L1M1 repeat: matches -1334. .-154 of consensus"
repeat_region 48335. .48727
/note="L1 repeat: matches 4030. .4226 of consensus"
repeat_region 48717. .49410
/note="L1M9 repeat: matches 5092. .5805 of consensus"
repeat_region 49411. .49506
/note="48 copies 2 mer at 69% conserved"
repeat_region 49528. .50023
/note="L1M9 repeat: matches 5787. .6308 of consensus"
repeat_region 50050. .50333
/note="AluY repeat: matches 1. .300 of consensus"

```

Query Match

7.48; Score 210.2; DB 68; Length 163253;

Best Local Similarity 60.1%; Pred. No. 1.4e-45;
Matches 385; Conservative 0; Mismatches 253; Indels 3; Gaps 2;

```
QY 428 acccttactctaccagtgtgagctctccgtggygcgacagcagtttaatggaagaaa 487
      ||||| ||| | ||| ||||| ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 159212 ACCTAAGATCTTTATGTTCAGCTCATGTAGGAAATAATGAATTTTGGGGGAAG 159271

QY 488 gatgagaccacccgtgaaacaacagatgccctgcccgtgcgtgaggaactctgcaga 547
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159272 CCAGACACAATGCTGCAATGAAGACATCCAGCACTGCAGATTGAACCTATTCCAG 159331

QY 548 acccctgcagaaaagtttgaagtaaatgaaagaaagcagagaagaaaacctcaata 607
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159332 GATCGCCTTAGAATGCTGAATCAGGAAGGA--TATAGATGATGACAAAGATGCA 159389

QY 608 atcgaataaagccaagtgttgaatltgcgtgaagcgaatttgcctgtgaatttga 667
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159390 GCCTGAGATCAGCTTAGTGTGTAATTCCTGAACCTGAAATATACCTGTAGTCTGA 159449

QY 668 ggtgccccggaggttgccccaccacacatgaagaacttgtgaccaggtttcaagttg 727
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159450 GGTATATTAAGAAAGTGAGCCACCACATATGAAAAGCTTGTACTCGAGTGTAGTAG 159509

QY 728 ggaatttgaaggaaggaaggaaggaagaaacaaagaatctccaaagaagaatgcg 787
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159510 AAAGTCTCTACAGAGAAGGAAGGAATAGCAAAAAAACTCTCCAAGAACGCTGCA 159569

QY 788 ggcgtctctggaagcagcttaagagcgtgcacacccctccctgctgtgagcgagtga 847
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159570 CACCGTCTTACAGAGGAGCTTAAAAAACTTCACCTCTTCTGTGTGGAAAAAGGCA 159629

QY 848 cagaatcaagaagaaaagtcagcccaacctgcaagacagccccgattatgccaagga 907
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159630 ATTTTAAAAAACGCCCTAAACAATAGTCAAGGCCGACCAAGATA-GGCCAAGGAT 159688

QY 908 gaatcctatagtagacttgcacagatccagcaggaacaaaggaaggaagcagagta 967
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159689 GAACCTATTTAGCTTCCTGGCGCAAAATTCAACAGGCCAAAAAGGAAGCAGATTA 159748

QY 968 catgctccttacagaacgaggtctccacgtcgcaggaagtttgcgtcaggttaaggt 1027
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159749 TGTTTTGTTCAGAAAGAGGAATGCCCTCGATGTCGAGAAATTTGTGATGTATGA 159808

QY 1028 tgggcatacacactgcagaaggaagtggttaccaataagaag 1068
      ||| ||| | |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 159809 AGGCAATGAAGCTGTACAGGAACGGGACCTAATAAAAAAGG 159849
```

Search completed: April 5, 2001, 09:36:07
Job time: 81412 sec

OM of: US-09-316-048-6 to: N_Geneseq_36:* out_format : pfs
Date: Apr 5, 2001 8:29 PM
About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09316048/runat_01042001_170039_7768/app_query.fasta_1.1293
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09316048_@CGN1_1_248 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:
Query: US-09-316-048-6
Query length: 577
Database: N_Geneseq_36:*
Database sequences: 480022
Database length: 187831343
Search time (sec): 885.780000

score_list:	Strd Orig	ZScore	EScore	Len	Documentation
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:293322		2970.00	4283.07	8.8e-231	3506
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:293324		2664.50	3841.00	3.7e-206	3142
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:293323		2659.00	3832.78	1.1e-205	3217
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X90786		2650.00	3819.81	5.6e-205	3190
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:293327		2633.50	3795.65	1.2e-203	3260
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:293328		2422.50	3490.90	1.2e-186	2859
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X39992		1880.00	2713.42	2.4e-143	1127
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X39992		1292.00	1864.27	4.7e-96	773
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X40000		1257.50	1813.32	3.3e-93	844
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X40002		1255.50	1809.25	5.5e-93	946
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:279981		697.00	1002.72	4.6e-48	660
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:280269		691.50	993.75	1.5e-47	727
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:217411		418.00	596.41	2.0e-25	769
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X24067		218.00	296.99	9.4e-09	1904
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:252512		211.50	290.73	2.1e-08	1396
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X03457		205.00	289.29	2.5e-08	1608
/SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:Q97228		157.00	195.66	0.0041	6671
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:T89799		157.00	195.66	0.0041	6671
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V27063		157.00	195.66	0.0041	6671
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:217757		147.00	212.90	0.0005	300
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:214226		147.00	212.90	0.0005	300
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:231402		133.50	174.06	0.0660	1969
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:234452		133.50	174.06	0.0660	1969
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:V73805		132.50	144.01	3.12	32207
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V19941		130.00	166.03	0.1850	2628
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:V01060		128.50	164.11	0.2365	28598
/SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:T03703		128.00	138.70	6.16	28598
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:T06769		128.00	138.57	6.26	28958
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:T89956		128.00	133.11	12.61	49377
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V05287		127.50	152.74	1.02	6755
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V15111		126.50	161.21	0.3431	2562
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:T59650		126.00	155.98	0.6714	3980
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:A07836		124.50	160.43	0.3792	2082
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:A05408		124.50	160.43	0.3792	2082
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:V52267		122.50	141.95	4.06	1265
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:207192		122.50	152.89	0.969	3275
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:A34721		122.00	146.32	2.32	6225
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:290111		122.00	150.61	1.34	3813

/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:290112	+	122.00	149.75	1.49	4148
/SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:Q43526	+	121.50	136.39	8.28	14255
/SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:Q75181	+	121.50	136.39	8.28	14255
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V20477	+	121.50	136.39	8.28	14255
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:292416	+	121.00	154.59	0.8023	2244

seq_name: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:293322

seq_documentation_block:

ID 293322 standard; cDNA; 3506 BP.

AC 293322;

DT 04-JUL-2000 (first entry)

XX Human staufen cDNA.

XX Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;

KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;

KW acquired immune deficiency syndrome; protease; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 409..2142

XX FT /*tag= a

XX FT /product= Staufen

XX CA2238656-A1.

XX 22-NOV-1999.

XX 22-MAY-1998; 98CA-2238656.

XX 22-MAY-1998; 98CA-2238656.

XX (UYMO-) UNIV MONTREAL.

XX Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;

XX Moulard AJ;

XX WPI; 2000-246924/22.

XX P-PSDB; Y83023.

XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful

PT for treating retroviral infections especially human immunodeficiency

PT virus infections (HIV)

XX Claim 4; Figure 1a; 96pp; English.

XX Staufen is a RNA binding protein which interacts with double stranded

CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA

CC transport and localization. Mammalian SP contain multiple ds

CC RNA-binding domains. Chimeric proteins comprising portions of

CC staufen with proteins with RNase or protease activity can be used to

CC treat virus infections. The RNase or protease activity of the fusion

CC protein prevents proper maturation of the virus. Mammalian staufen

CC proteins seem to recognise double stranded RNA structure rather than

CC any sequence specific position. 2 bacterially-expressed fusion

CC proteins used in an RNA-binding assay (his/SP (human SP) and

CC MBP/SP (murine SP) both proteins strongly bound double stranded

CC RNA. Both fusion proteins also directly bound labeled ds RNAs and

CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the

CC treatment of RNA virus infections, especially those caused by

CC retroviruses, in particular human immunodeficiency virus.

XX Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

alignment_scores:

Quality: 2970.00

Ratio: 5.147

Percent Similarity: 100.000

Length: 577

Gaps: 0

Percent Identity: 100.000

alignment_block:
US-09-316-048-6 x 293322 ..

Align seg 1/1 to: 293322 from: 1 to: 3506

```
1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG1 17
|||||
409 ATGTCACAGTTCACAGTGCAGATTCAGAACCCATCTGCTGCTCTCAGG 458
17 ySerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeuMets 34
|||||
459 GAGCCAATACTGACACAGAACCGATCTCTCTCTCAGAGCCTTGATGA 508
34 erIleProSerThrThrSerSerLeuProSerGluAsnAlaGlyArgPro 50
|||||
509 GTATTCTTCTACTACTAGCTCTCTGCCCTCTGAAATGACAGGTAGACCC 558
51 ILeGlnAsnSerAlaLeuProSerAlaSerIleThrSerThrSerAlaAl 67
|||||
559 ATTCAAACCTGCTTTACCTCTGCATCTATTACATCCACACAGTCAGC 608
67 alaGluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysL 84
|||||
609 TGCAGAAAGCATAACCCCTACTGTAGAACATAATGCACCTGTCATGAAC 658
84 euGlyLysLysProMetTyrLysProValAspProTyrSerArgMetGln 100
|||||
659 TTGGAATAAACCAATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAG 708
101 SerThrTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProArgTyrPh 117
|||||
709 TCCACCTATACTACACATGAGAGGAGGTCTTATCCCCGAGGACTT 758
117 eTyrProPheProValProProLeuLeuTyrGlnValGluLeuSerValG 134
|||||
759 TTACCCATTTCAGTTCACCTTACTTTATCAAGTGGAACTTTCTGTGG 808
134 IyGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 150
|||||
809 GAGGACAGCAATTATAAGCAAGAAAGAAAGACAGAGCTGCGAAACAC 858
151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProG1 167
|||||
859 GATGCTGCTGCCAAAGCGTTGAGGATCCGACAGATGAGCCCTGCCAGA 908
167 uArgLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysS 184
|||||
909 GAGGCTGGAGGTGAATGGAAGAAATCCGAAGAAGAAATCTCAATAAAT 958
184 erGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProVal 200
|||||
959 CTGAATTAAGTCAAGTGTGGAGATTGCACTTAAACGGAACCTGCCGTG 1008
201 AsnPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheVa 217
|||||
1009 AATTTCGAGGTGGCCGGGAGAGTGGCCACCCACATGAAGAACTTTGT 1058
217 lThrLysValSerValGlyGluPheValGlyGluGlyGlyLysSerL 234
|||||
1059 GACCAAGGTTTCGTTGGGGAGTTGTGGGGGAAAGGTGAAGGGGAAAGCA 1108
234 yLysIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLys 250
|||||
1109 AGAAGATTTCAAGAAATGCGCCCATAGCTGTCTTGAGGAGCTGAAG 1158
251 LysLeuProProLeuProAlaValGluArgValLysProArgIleLysLy 267
|||||
1159 AAGTTACCGCCCTGCTGAGTTGAACGAGTAAAGCCTAGAAATCAAAA 1208
267 sLySThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnG 284
|||||
1209 GAAACAAACCAATAGTCAAGCCACAGACAGCCCAAGAATATGGCCAGG 1258
```

```
284 lYlLeAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlu 300
|||||
1259 GGATCAATCCGATTAGCCGACTGGCCAGATCCAGACAGGCAAAAAAGAG 1308
301 LysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgAr 317
|||||
1309 AAGGAGCCAGAGTACAGCTCTCAGAGAGGAGGCTCCCGCCGCGCAG 1358
317 gGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrg 334
|||||
1359 GGAGTTTGTGATGCAAGGTGAAGTTGGAACCAACACTGCAGAGAACGG 1408
334 lYThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 350
|||||
1409 GCACCAACAAGAGGTGGCCAAAGCGCAATGCAGCCGAGAACATGCTGGAG 1458
351 lLeuGluPheLysValProGlnArgGlnProThrLysProAlaLeuLy 367
|||||
1459 ATCCTTGTTTCAAGTCCCGCAGCGGCGAGCCCAACCAACCCGACTCAA 1508
367 sSerGluGluLysThrProIleLysLysProGlyAspGlyArgLysValT 384
|||||
1509 GTCAGAGAGAGAACACCCATTAAGAAACCAAGGGGATGGAAGAAAGTAA 1558
384 hrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGlu 400
|||||
1559 CCTTTTGAACCTGCTCTGGGATGAATAAGGACTAGTAATATAAGAG 1608
401 AspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyI1 417
|||||
1609 GATGAGTTCAAGATGCTTATCTAAGTCATGACGACGCTGCTGTAAT 1658
417 eleuProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyH 434
|||||
1659 TCTTCCCATGTGTCGGAGGTGCGCCAGGCTGTAGAGTTAGTCAAGGAC 1708
434 lSHISThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr 450
|||||
1709 ATCACACCAAGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCAG 1758
451 ValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProTh 467
|||||
1759 GTAACGTCATGATAGCCCGAGAGTTGTATGGGGCACCCTGCCCTCAC 1808
467 rAlaGluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisG 484
|||||
1809 AGCCGAGACCATTTTAAAGAATAACATCTCTTCAGGCCACGTACCCCAT 1858
484 lYProLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGln 500
|||||
1859 GACCTCTCACGAGACCTCTGAGCAACTGACTATCTTCCAGAGTCCAG 1908
501 GlyPheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPh 517
|||||
1909 GGATTCCAGGTTGAATACAAAGACTTCCCAAAAACAACAAGAACGAAT 1958
517 eValSerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyI 534
|||||
1959 TGTATCTCTTATCAATTGCTCTCTCAGCCACCTGTGATCAGCCATGTA 2008
534 lGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeu 550
|||||
2009 TCGGCAAGATGTGAGTCCCTGCCATGATATGGCTGCGCTGAACATCTTA 2058
551 LysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrgL 567
|||||
2059 AAGTTGCTGCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAGG 2108
567 yAsnGlyProMetSerValCysGlyArgCys 577
|||||
2109 AAACGGACCAATGTCTGTGTGTGGAGGTGC 2139
```



```

171 TTTTTCCTTCTTCCTTCCC..TCCTCGCCGCCACCGCCAGGACCGCG 219
48 IYArg.....ProIleGlnAsn.....SerAlaLeuPro 57
   |||||      |||  ||:::  |||
220 GCCGGGGCAGAGTCCGGAGCAGCACGACGATTTATTAACTTAACC 269
   |||||      |||  ||:::  |||
58 SerAlaSerIleThrSerThrSerAla.....Al 67
   |||  :::  ||:::||||:::
270 TCTCAGACTGAACAAGAAGACACATGTGTTCTCTGGAACGCCCTCTTTTAA 319
67 aaIagLserIleThrProThrValGIuleuAsnAlaleuCysMetLysL 84
   |||||      |||||      |||||      |||||      |||||      |||||
320 AAAAGAAAGCATAAACCCCTACTGTAGAACTAAATGCACGTGCATGAAC 369
84 euGLyLSysLPrometTyrrLysProvalAspproTyrrSerArgetIn 100
   |||||      |||||      |||||      |||||      |||||      |||||
370 TTGAAAAAACAACCAATGTATAAGCCTGTGACCCCTTACTCTCGATGCAG 419
101 SerThrTyrrAsnTyrrAsnMetArGLylalatyrrProPArgTyrrPh 117
   |||||      |||||      |||||      |||||      |||||      |||||
420 TCCACCTATACTACACACATGAGAGGAGTGCTTATCCCCCGAGGTACTT 469
117 eTYrrProPheProvalProProleuLeuTYrrGlnValGIuleuSerValG 134
   |||||      |||||      |||||      |||||      |||||      |||||
470 TTACCCAATTCCAGTTCACCTTTACTTTATCAAGTGAACCTTCTGTGG 519
134 lYglYnglInPheasnGlYlysGlYlystrArGlnAlaAlaLysHis 150
   |||||      |||||      |||||      |||||      |||||      |||||
520 GAGGACAGCAATTTAATGGCAAAGAAAAGACACAGCAGGCTGCGAAACAC 569
151 AspAlaAlaAlaLysAlaLeuArGIleleuGlnAsnGIuProleuProGI 167
   |||||      |||||      |||||      |||||      |||||      |||||
570 GATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGATGAGCCCCCTGCCAGA 619
167 uArGLeuGIuValAsnGlYarGLuSerGIuGIuGIuAsnLeuAsnLysS 184
   |||||      |||||      |||||      |||||      |||||      |||||
620 GAGGCTGGAGGTGAATGGAAGAGAATCCGAAGAAGAAAATCTCAATAAAT 669
184 erGLuIeSerGIuValPheGIuIleAlaleuLysArGsnLeuProval 200
   |||||      |||||      |||||      |||||      |||||      |||||
670 CTGAATTAAGTCAAGTGTGTGAGATTGCACCTTAAACGGAACCTTGCTGTG 719
201 AsnPheGIuValAlaArGLuSerGIuProProHisMetLysAsnPheVa 217
   |||||      |||||      |||||      |||||      |||||      |||||
720 AATTTCGAGGTGGCCCCGGAGAGTGCCCCACCCACATGAAGAACTTTGT 769
217 lThrLysValSerValGIylGIuPheValGIylGIuGIylGIylLysSerL 234
   |||||      |||||      |||||      |||||      |||||      |||||
770 GACCAAGTTTCGGTTGGGGAGTTTGTGGGGAGGTGAAGGGAAGCA 819
234 ySLysIeSerLysLysAsnAlaAlaIleAlaValleuGIuGIuLeuLys 250
   |||||      |||||      |||||      |||||      |||||      |||||
820 AGAAGATTTCAAAGAAAAATGCCCATAGCTGTCTTGAGGAGCTGAAG 869
251 LysLeuProProleuProAlaValGIuArGVAlLysProArGIleLysLy 267
   |||||      |||||      |||||      |||||      |||||      |||||
870 AAGTTACGCCCCCTGCCTGCAGTTGAACGAGTAAGCCTTAGAATCAAAAA 919
267 sLYstHrLysProIleValLysProGIInThrSerProGIuTYrrGIyng 284
   |||||      |||||      |||||      |||||      |||||      |||||
920 GAAAAAACAACCATAGTCAAGCCACAGACAAGCCAGAATATGGCCAGG 969
284 lyIleasNProlIeSerArGVleuAlaglnIlegInAlaLysLysGIu 300
   |||||      |||||      |||||      |||||      |||||      |||||
970 GGATCAATCCGATTAGCCGACGTGGCCAGATCCACAGGCAAGAAAAAGAG 1019
301 LysGIuProGIuTYrrThrLeuDeuThrGIuArGIyLeuProArGVArGa 317
   |||||      |||||      |||||      |||||      |||||      |||||
1020 AAGGAGCCAGAGTACACGCTCTCACAGAGCGAGGCTCCCGCGCGCAG 1069
317 gGIuPheValMetGIuValLysValGIylAsnHisThrAlaGIuGIyThrg 334
   |||||      |||||      |||||      |||||      |||||      |||||
1070 GGAGTTTGTGATGCAGGTGAAGGTTGGAAACCACTGTCAGAAAGGAAACGG 1119

```

334	lyThAsnLysLysValAlaIAsnArgAsnAlaAlaGluAsnMetLeuGlu	350
1120	GCACCAACAAGAGGTGGCCAAAGCCAAATGCAGCCGAGAACATCTGGAG	1169
351	lleuGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLys	367
1170	ATCCTTGGTTTCAAGTCCCGCAGGGCCAGCCCAACCAACCCGCACTCAA	1219
367	sSerGluGluLysThrProIleLysLysProGlyAspGlyArgLysValT	384
1220	GTCAGAGGAGAAGACACCCCATAAAGAAACCAGGGGATGGAAGAAAGTAA	1269
384	hrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGlu	400
1270	CCTTTTTGAACCTGGCTCTGGGATGAAAATGGGACTAGTATTAAGAG	1319
401	AspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyI	417
1320	GATGAGTTTCAGGATGCCTTATCTTAAGTCATCAGCAGCTGCCCTGCGAAT	1369
417	eleuProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyH	434
1370	TCCTCCCATGGTGCCCGAGGTGCCCCAGGCTGTAGAGTTAGTCAAGGAC	1419
434	ISHISThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr	450
1420	ATCAACAACAAGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACG	1469
451	ValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProTh	467
1470	GTAAC TGCCATGATAGCCCCGAGAGTTGTGTATGGGGCACCCTCGCCAC	1519
467	raIaGluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisG	484
1520	AGCCGAGACCATTTTAAAGATATACATCTCTTCAGGCCACGTACCCCATG	1569
484	lyProLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGln	500
1570	GACCTCTCAGCAGACCCCTGTGAGCACTGACTATCTTCCAGAGTCCAG	1619
501	GlyPheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPh	517
1620	GGATTCCAGGTGTAATACAAAGACTTCCCAAAAAACAACAAGACGAATT	1669
517	evaSerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyI	534
1670	TGTATCTCTTATCAATGTGCTCTCTCAGCCACCTCTGATCAGCCATGGTA	1719
534	IeglyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeu	550
1720	TGGGCAAGGATGTGAGTCTCTGCCATGATATGGCTGCGCTGAACATCTTA	1769
551	LysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGln	567
1770	AAGTTGCTGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGACAGG	1819
567	YAsnGlyProMetSerValCysGlyArgCys	577
1820	AAACGAGCAATGCTCTGTGTGTGGGAGGTGC	1850

```
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X90786
seq_documentation_block:
ID   X90786 standard; DNA; 3190 BP.
XX
AC   X90786;
XX
DT   13-JAN-2000 (first entry)
XX
DE   DNA encoding human staufen protein.
XX
KW   Human staufen protein; hstau; protein lysate; lung; kidney; testis;
```

KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;
telomerase assembly; maturation; transport; regulation; interaction;
inhibitory agent; antibody; immunoassay; ds.
XX Homo sapiens.

XX Key Location/Qualifiers
FH CDS 345..1835
FT /*tag= a
FT /product= "Human staufen protein"
FT /note= "Binds to telomerase RNA"
FT polyA_signal 3172..3177
FT /*tag= b

PN W09951255-A1.
XX 14-OCT-1999.
XX
PF 06-APR-1999; 99WO-US07533.
XX
XX 06-APR-1998; 98US-0080783.
XX

XX (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
XX Greider CW, Le S;
XX
XX WPI: 1999-620168/53.
XX P-PSDB; Y26335.
XX

PT Human staufen polypeptide useful in methods for identifying telomerase
PT inhibitors
XX
PS Clalm 13; Page 41-42; 50pp; English.

XX The present sequence is an isolated DNA encoding human staufen (hstau)
CC protein. The hstau protein was detected in protein lysates from lung,
CC kidney, testis and ovary, but not in brain and heart. It binds
CC specifically to human telomerase RNA and can complex with the
CC telomerase catalytic subunit. The hstau protein may play a role in
CC telomerase assembly, maturation, transport and regulation. Interaction
CC between hstau and telomerase can be inhibited by administering inhibitory
CC agents to the cell. These telomerase inhibitors can be identified using
CC hstau protein. Antibodies generated against hstau can be used in various
CC immunoassays.
XX

SQ Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;

alignment_scores:
Quality: 2650.00 Length: 577
Ratio: 4.907 Gaps: 2
Percent Similarity: 93.588 Percent Identity: 90.815

alignment_block:
US-09-316-048-6 x X90786 ..

Align seg 1/1 to: X90786 from: 1 to: 3190

15 LeuSerGlySerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnPr 31
103 CTCACGGCCACATCCGCTCTTCCCTCCCTCCCTCGTCCCTCTTCCCTCC 152
31 OleuMetSerIleProSerThrThrSerSerLeuProSerGluAsnIag 48
153 TTTTTCCTTCTTCCCTTCCCTTCCCTCCCTCCCTCCCTCCCTCCCTCC 201
48 lYArgProIleGlnAsnSerAlaLeuPro..... 57
202 GCCGGGGGAGAGCTCGAGCAGCAGCAGCAGCTTTATTACCATTAACC 251
58 SerAlaSerIleThrSerThrSerAla.....Al 67
252 TCTCAGAACTGAACAAGACACATTTGTTCTGGAACGCCCTCTTTTAA 301

67 aAlaGluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysL 84
302 AAAAGAAAGCATTAACCCCTACTGTAGACTTAATGCACCTGTGCATGAAC 351
84 eUGlyLysLysProMetTyrLysProValAspProTyrSerArgMetGln 100
352 TTGGAATAAAACCAATGTATAGCCCTGTTGACCCCTTACTCTCGCATGCCG 401
101 SerThrTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProArgTyrPh 117
402 TCCACCTATACTACACATGAGAGGAGGTGCTTATCCCCCGAGGTACTT 451
117 eTyrProPheProValProProLeuLeuTyrGlnValGluLeuSerValG 134
452 TTACCCATTTCAGTCCAGTCCACCTTACTTTATCAAGTGGAACCTTCTGTCG 501
134 lYgLyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 150
502 GAGGACAGCAATTATATGGCAAGGAAGAGACAGACAGAGCTGCGAAACAC 551
151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGl 167
552 GATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGATGAGCCCTGCCACA 601
167 uArgLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysS 184
602 GAGGCTGGAGGTGAATGGAAGAGAATCCGAGAGAAATAATCTCAATTAAT 651
184 eRgIuIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProVal 200
652 CTGAATAATAGTCAAGTCTTTGAGATTGACATTAAACGGAACCTTGCTGTG 701
201 AsnPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheVa 217
702 AATTTGAGGTGGCCGGGAGAGAGTGCCACCCACATGAAGAATTTGT 751
217 lThrLysValSerValGlyGluPheValGlyGluGlyGlyLysSerL 234
752 GACCAAGGTTCCGTTGGGGAGTTGTGGGGGAAGGTGAAGGAAAGCA 801
234 ySLysIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLys 250
802 AGAAGATTTCAAAGAAATAATGCCCGCATAGCTGTCTTGAGGAGCTGAAG 851
251 LysLeuProProLeuProAlaValGluArgValLysProArgIleLysLy 267
852 AAGTTACCGCCCTGCTGCTGAGTTGAACGAGTAAAGCCTAGAAATCAAAA 901
267 sLysThrLysProIleValLysProGlnThrSerProGluTyrGlyGln 284
902 GAAACCAAAACCCATAGTCAAGCCACAGACAGACCCAGAAATATGCCAGG 951
284 lYIleAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlu 300
952 GGATCAATCCGATAGCCGAGCTGGCCAGATCCAGACGCAAAAAAGAG 1001
301 LysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgAr 317
1002 AAGGAGCCAGAGTACAGCTCTCTCACAGACGAGGCCCTCCCGCCGACAG 1051
317 gGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrg 334
1052 GGAGTTTGTGATGACAGGTGAAGTTGGAACACACACTGCAGAAAGAACGG 1101
334 lYThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 350
1102 GCACCAACAAGAGGTGGCCAAAGCGCAATGCAGCCGAGAACATGCTGAG 1151
351 lIleuGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLy 367
1152 ATCCTTGGTTTCAAGTCCCGCAGGCGCAGCCACCAACCCGCACTCAA 1201

367 sSerGlulGlulysThrProIleLysLysProGlyAspGlyArgLysValT 384
|||||
1202 GTCAGAGGAGAGACACCCATAAGAAACCAGGGATGGAAAGAAAGTAA 1251
384 hrPhePheGluProGlySerGlyAspGlyuasnGlyThrSerAsnLysGlu 400
|||||
1252 CCTTTTGAACCTGGCTCTGGGGATGAAAAATGGACTAGTAATAAGAG 1301
401 AspGluPheArgMetProGlyrLeuSerHisGlnGlnLeuProAlaGlyI 417
|||||
1302 GATGAGTTCAGGATGCCCTATCTAAGTCATCAGCAGCTGCTGCGAAT 1351
417 eLeuProMetValProGluValAlaGlnAlaValGlyValSerGlnGly 434
|||||
1352 TCTTCCCATGGTGCCCGAGGTCCGCCAGGCTGTAGAGATTAGTCAAGGAC 1401
434 iShiThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr 450
|||||
1402 ATCACACCAAGATTTTACCAAGGGCAGCTCCGAATCTGCCAAGGCCACG 1451
451 ValThrAlaMetIleAlaArgGluLeuLeuTyrglyThrSerProth 467
|||||
1452 GTACTGCCATGATAGCCGAGAGTTGTTGATGGGGCCACCTCGCCAC 1501
467 rAlaGluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisG 484
|||||
1502 AGCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCACGTACCCCATG 1551
484 lProLeuThrArgProSerGluGlnLeuAspTyrlLeuSerArgValGln 500
|||||
1552 GACCTCTCAGAGACCCCTTGAGCAACTGGACTATCTTCCAGAGTCCAG 1601
501 GlyPheGlnValGluTyrlLysAspPheProLysAsnAsnLysAsnGluPh 517
|||||
1602 GGATTTCCAGGTTGAATACAAAGACTTCCCCAAAAACAACAAAGCAATP 1651
517 eValSerLeuIleAsnCysserSerGlnProProLeuIleSerHisGlyI 534
|||||
1652 TGTATCTCTTATCAATGTGCTCTCAGCCACCTGTGATCAGCCATGCTA 1701
534 lGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeu 550
|||||
1702 TCGGCAAGGATGTGAGTCTGCCATGATATGCTGCGCTGAACATCTTA 1751
551 LysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrG 567
|||||
1752 AAGTTGCTGCTGAGTGTGACCAACAAGTACAGAGATGCCAAGAACAGG 1801
567 yAsnGlyProMetSerValCysGlyArgCys 577
|||||
1802 AAACGGACCAATGTGTGTGTGGAGGTGC 1832

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:293327

seq_documentation_block:
ID 293327 standard; cDNA; 3260 BP.

XX 293327;
AC
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauften cDNA.
XX
KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; ; human; ss.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 409..1311
FT /*tag= a
FT /product= Staufen protein

XX CA2238656-A1.
PN
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83108.

PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)

PS Claim 4; Figure 1b; 96pp; English.

XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hSP (human SP) and
CC MBP/mSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.

SQ Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

alignment_scores:
Quality: 2633.50 Length: 509
Ratio: 5.184 Gaps: 1
Percent Similarity: 99.804 Percent Identity: 99.804

alignment_block:
US-09-316-048-6 x 293327 ..

Align seg 1/1 to: 293327 from: 1 to: 3260

69 GluSerIleThrProThrValGluLeuAsnAlaLeucysmetLysLeuG 85
|||||
370 GAAAGCATAAACCCCTACTGTAGAACTAATGCACGTGTCATGAACCTTG 419
85 yLysLysProMetTyrlLysProValAspProTyrlSerArgMetGlnSert 102
|||||
420 AAAAAAACCAATGTATAAGCCTGTTGACCCTTACTCTCGGATGCACTCC. 468
102 hrTyrlAsnTyrlAsnMetArgGlyGlyAlaTyrlProProArgTyrlPheTy 118
|||||
469 .TATTACTACACATGAGAGAGGTGCTTATCCCGAGGTACTTTTAC 516
119 PropheProValProProLeuLeuTyrglnValGluLeuSerValGlyG 135
|||||
517 CCATTTCCAGTTCACCTTTACTTTATCAAGTGAACCTTCTGTGGGAGG 566
135 yGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAsp 152
|||||
567 ACAGCAATTTAATGGCAAGAAAGAACAGACAGGCTGCGAAACACAGATG 616

152	laAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArg	168
617	CTGCTGCCAAAGCGTTGAGGATCTTCAGAAATGAGCCCCCTGCCAGAGAGG	666
169	LeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSerG1	185
667	CTGGAGGTGAATGGAAGAGAAATCCGAAGAAGAAATCTCAATAAATCTGA	716
185	uIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsnP	202
717	AATAAGTCAAGTGTTCAGATTGCACCTTAACGGAACTTGCCCTGTGAATT	766
202	heGluValAlaArgGluSerGlyProProHisMetLysAsnPheValThr	218
767	TCGAGGTGGCCCGGAGAGTGGCCCAACCACATGAAGAACTTTGTGACC	816
219	LysValSerValGlyGluPheValGlyGluGlyGlyLysSerLysLy	235
817	AAGTTTCGGTTGGGAGTTTGGGGGAAGGTGAAGGGAAGAAACAAGAA	866
235	sIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLysLysL	252
867	GATTTCAAAAGAAAAATGCCGCCATAGCTGTCTTGAGAGAGCTGAAGAAGT	916
252	euProProLeuProAlaValGluArgValLysProArgIleLysLysLys	268
917	TACCGCCCCCTGCCTGCAGTTGACGAGTAAGAAGCTAGATCAAAAAGAAA	966
269	ThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnGlyI1	285
967	ACAAAAACCATAGTCAAGCCACAGACAAAGCCAGAAATATGCGCCAGGGGAT	1016
285	eAsnProIleSerArgLeuAlaGlnIleGlnAlaLysLysGluLysG	302
1017	CAATCCGATTAGCCGACTGGCCAGATCCAGCAGAGCAAAAAAGAGAGAAGC	1066
302	LuProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgGlu	318
1067	AGCCAGAGTACACGCTCTCTACAGAGAGGAGGCTCCCGCGCCGAGGAG	1116
319	PheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyTh	335
1117	TTTGTGATGCAGGTGAAGGTTCGAACACACACTGCAGAGAAGAACGGGCAC	1166
335	rAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleL	352
1167	CAACAAGAAAGTGGCCCAAGCCCAATGCAGCCGAGAACATGCTGCAGATCC	1216
352	eugLysPheLysValProGlnArgGlnProThrLysProAlaLeuLysSer	368
1217	TTGGTTTCAAAAGTCCCGCAGCGCAGCCACCAAAACCCGACTCAAGTCA	1266
369	GluGluLysThrProIleLysLysProGlyAspGlyArgLysValThrPh	385
1267	GAGGAGAAGACACCCATAAAGAACACAGGGGATGGAAGAAAGTAACTT	1316
385	ePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspG	402
1317	TTTGTGAACCTGGCTCTGGGGATGAAAAATGGGACTAGTAATAAAGAGATG	1366
402	LuPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeu	418
1367	AGTTCAAGGATGCCTTATCTAAGTCATCAGCAGCTGCCTGTGGAATTCTT	1416
419	PrometValProGluValAlaGlnAlaValGlyValSerGlnGlyHisH1	435
1417	CCCATGGTGCCCGAGGTGCGCCAGGCTGTAGAGAGTTAGTCAAGACATCA	1466
435	sThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValT	452
1467	CACCAAAAGATTTTACCAGGCGAGCTCCGAATCTCTGCCAAAGCCACGGTAA	1516

452 hralametiIleaIaargGluleuLeuTyrglyGlyThrSerProthraIa 468
 1517 CTGCCATGATAGCCCGAGAGTGTGTATG6GGGGCACCCTGCCACAGCC 1566
 469 GluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisGlyPr 485
 1567 GAGACCATTTTAAAGAATAACATCTTTCAGGCCACGTACCCTCATGACC 1616
 485 oleuThrArgProSerGluGlnLeuAspTyrlleuSerArgValGlnGlyP 502
 1617 TCTCAGCAGACCCCTGTGAGCAACTGCACTATCTTCCAGAGTCCAGGAT 1666
 502 heGlnValGluTyrlLysAspPheProLysAsnAsnLysAsnGluPheVal 518
 1667 TCCAGGTGGAATACAAAGACTTCCCAAATAACAACAAGACGAATTGTA 1716
 519 SerIleuIleAsnCysSerSerGlnProProIleuIleSerHisGlyIleGl 535
 1717 TCTCTTATCATTTGCTCTCCTCCTCAGCCACCCTGTGATCAGCCATGGTACGG 1766
 535 yLysAspValGluSerCysHisAspMetAlaIaIaLeuAsnIleLeuLysL 552
 1767 CAAGCATGTGGAGTCTCCATCATATAGCTGCGCTGAACATCTTAAAGT 1816
 552 euleuSerGluLeuAspGlnGlnInsertThrGluMetProArgThrGlyAsn 568
 1817 TGTGTCTGTAGTTGGACCAACAAGTACAGAGATGCCAAGAACAGGAAC 1866
 569 GlyPrometSerValCysGlyArgCys 577
 1867 GACCAATGCTGTGTGTGGAGGTGC 1893

```
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:293328
```

```
seq_documentation_block:
ID 293328 standard; cDNA; 2859 BP.
```

DT	04-JUL-2000	(first entry)
XX		
AC	293328;	
AA		

XX Mouse staußen cDNA.
DE
XX
XX Staußen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; mouse; ss.
XX
OS Mus musculus.

Key	Location/Qualifiers
FT CDS	325..1788
FT	/*tag= a
FT	/product= Stauflen protein

PN CA2238656-A1.

PD 22-NOV-1999.

PF 22-MAY-1998; 98CA-2238656.

PR 22-MAY-1998; 98CA-2238656.

.PA (UYMO-) UNIV MONTREAL.

PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;

PI Mouland AJ;

DR WPI; 2000-246924/22.

DR P-PSDB; Y83024.

XX

PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)

XX Claim 4; Figure 1c; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;

alignment_scores:
Quality: 2422.50 Length: 552
Ratio: 4.750 Gaps: 2
Percent Similarity: 92.391 Percent Identity: 85.145

alignment_block:
US-09-316-048-6 x 293328 ..

Align seg 1/1 to: 293328 from: 1 to: 2859

26 SerLeuLeuSerGlnProLeuMetSerIleProSerThrThrSerSerle 42
|||||
157 TCCTCTCTGCTGCTGGGCCCTTTTCTCCGCTCTTCACTTGCTTCTTC 206
42 uProSerGluAsnAlaGlyArgProIleGlnAsnSerAlaLeuProSera 59
|||
207 ACCTCTCGCCGCCGCCAAGA.....CCGCCGG 235
59 IAserIleThrSerThrSerAlaAlaAlaGluSerIleThrProThrVal 75
||
236 CCCCCGGAGAGCTCTGGGGAAGACAGCCAGAAAGTATAGCTTCTACCAATT 285
76 GluLeuAsnAlaLeuGlyMetLysLeuGlyLysLysProMetTyrlsPr 92
|||||
286 GAGCTCAATGACACTGTGTGTGAACCTGGAAGAAACAATGTATTAAGCC 335
92 oValAspProTyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArg 109
|||||
336 CGTGACCCCTCACTCTCGGATGACATCCACTACGCTATGCGATGCGTG 385
109 lYgIyAlaTyrProProArgTyrPheTyrProPheProValProProLeu 125
|||||
386 GAGGTGCTATCCGCCAGATACTTTTACCACATTCTCAGTCCACCTTTA 435
126 LeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPheAsnGlyTysG 142
|||||
436 CTCTACCAAGTTGAGCTCTCCGTGGCGGACAGCAGTTTAAATGGAAAG 485
142 yLysThrArgGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuArg 159
|||
486 AAGATGAGACACCCCGTGAACACGATGCCCTGCGCGTGAGGA 535
159 lLeuGlnAsnGluProLeuProGluArgLeuGluValAsnGlyArgGlu 175
|||||
536 CTCTGCAGAGTGAACCCCTGCCAGAAAGTTGAGGTAATGGAAGAGAA 585
176 SerGluGluGluAsnLeuAsnLysSerGluLysSerGlnValPheGlu 192
:::
586 GCAGAGGAGAAAACTCAATAATAATCGGAATTAAGCCAAAGTGTGAAAT 635
192 eAlaLeuLysArgAsnLeuProValAsnPheGluValAlaArgGluSer 209

636 TCGCCTGAAGCGGAATTCCTGTGAATTTTGAAGTGGCCCGGAGACTG 685
209 lYProProHisMetLysAsnPheValThrLysValSerValGlyGluPhe 225
|||||
686 GCCCACCACACATGAAGACTTTGTGACCAGCGGTTTCAGTTGGGAATTT 735
226 ValGlyGluGlyGlyLysSerLysLysLysLysLysLysLysAsnAla 242
|||||
736 GTAGGGGAGAGAGAGGAAAGCAAGAGATCTCCAAGAAATGCGGC 785
242 AlAlaValLeuGluLeuLysLysLysLeuProProLeuProAlaValG 259
|
786 CAGGCTGTCTTGAGAGACTTAGAGGCTGCCACCCCTCGCTGTGG 835
259 lUArgValLysProArgLysLysLysLysLysThrLysProLysValLysPro 275
|||||
836 AGCAGTGAAGCCAGATCAAGAAAGTCAAGCCACCTGCAAG... 882
276 GlnThrSerProGluTyrGlyGlnGlyLysLysLysProIleSerArgLeu 292
|||
883 ..ACAGCCCCGATTTATGCCCAAGGATGAATCTTATAGTACTTGC 929
292 aGlnIleGlnAlaLysLysGluLysGluProGluTyrThrLeuLeu 309
|||||
930 ACAGATCCAGCAGCAAAAGAGAGAGAGCCAGATACATGCTCTCTTA 979
309 hrcGluArgGlyLeuProArgArgArgGluPheValMetGlnValLysVal 325
|||||
980 CAGAACGAGGCTCTCCACGTCGACAGGAGTTTGATGACAGTAAAGCTT 1029
326 GlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLysValAlaLysAr 342
|||
1030 GGGCATCACACTGCAGAGAGAGTGGTACCAATAGAAGGTGGCCAAGCG 1079
342 gAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysValProGlnA 359
|||||
1080 TAATGCTGCTGAGAACATGCTGAGATCTCGGGTTCAAGTTCCCCAGG 1129
359 rGlnProThrLysProAlaLeuLysSerGluGluLysThrProIleLys 375
|||
1130 CGCAGCCTGCCAAGCCAGCAGCTCAATCAGAGAGAGACTCCAGTAAAG 1179
376 LysProGlyLysGlyArgLysValThrPhePheGluProGlySerGlyAs 392
1180 AAACCAAGAGAGCGAAGGAAGTAACGTTTGTGAACCTAGCCCTGGGA 1229
392 pGluAsnGlyThrSerAsnLysGluAspGluPheArgMetProTyrLeu 409
|||||
1230 TGAATAATGGAACCTAGTAACAAGAGAGAGAGTTCAGATGCCTTATCTTA 1279
409 eRhISgInGlnLeuProAlaGlyIleLeuProMetValProGluValAla 425
|||||
1280 GCCATCAGCAGCTGCCAGCTGGAATTCTCCCATGCTGCCGGAAGTTGCC 1329
426 GlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArgAl 442
|||||
1330 CAGGCTGTGGGGTTAGTCAAGGACACACACCAAGATTTCACACAGGCG 1379
442 AlAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 459
|||||
1380 AGCTCCAAATCTGCAAGCAACGTAAGTCCATGATAGCCCGAGAGT 1429
459 eUeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsn 475
|||||
1430 TGTGTACGGGGGACCTCGCCACAGCCGAGACCAATTTAAAGATTAAC 1479
476 lIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 492
|||||
1480 ATCTCTTACAGCCACGTAACCATGACCTCGCACTAGACCTCTGAGCA 1529
492 nLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAsp 509
|||

1530 ACTGTACTACCTTTCCAGAGCCAGGATTCACAGTTGAATACAAGATT 1579
509 heprolysAsnAsnLysAsnGluPheValSerLeuIleAsnCysSerSer 525
|||||
1580 TTCCCAAGAACAACAAGAACGAGTGTATCTCTCATCAACTGCTCTCA 1629
526 GluProProLeuIleSerHisGlyIleGlyLysAspValGluSerCysHis 542
|||||
1630 CAGCCGCTCTCTGTCATGTCATGCGCAAGGATGTGAGTCTGTCA 1679
542 saspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLeuAspGlnG 559
|||||
1680 TGATATGGCTGCACATGAACATTTAAAGCTGCTGTGAGTTGACCAAC 1729
559 InSerThrGluMetProArgThrGlyAsnGlyPrometSerValCysGly 575
|||||
1730 AGAGCACAGAGATGCCAAGAACAAGAAATGACCAAGTTTCAGCGTGGGG 1779
576 ArgCys 577
|||||
1780 AGGTGC 1785

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X39991

seq_documentation_block:

ID X39991 standard; DNA; 1127 BP.

XX X39991;

DT 02-JUL-1999 (first entry)

DE Prostate cancer associated gene.

KW Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer; ss.

OS Homo sapiens.

PN WO9904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

DR WPL; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 625-626; 787pp; English.

CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

SQ Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;

alignment_scores:

Quality: 1880.00 Length: 375
Ratio: 5.054 Gaps: 1
Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:

US-09-316-048-6 x X39991 ..

Align seg 1/1 to: X39991 from: 1 to: 1127

135 GlyGlnGlnPheAsnGlyLysGlyThrArgGlnAlaAlaLysHisAs 151

|||||
3 GGACAGCAATTTATGGCAAGGAAGAACAGACAGAGCTGCGAAACACGA 52

151 PALAAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluA 168

|||||
53 TGCTGCTGCCAAGCGTTGAGGATCTGCAGAAATGAGCCCTGCCACAGAA 102

168 rGLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSer 184

|||||
103 GGCTGGAGGTGAATGGAAGAAATCCGAAAGAAAGAAATCTCAATAATCT 152

185 GluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAs 201

|||||
153 GAAATAAGTCAAGTGTGAGATTGCACTTAAACGGAACCTGCTGTGAA 202

201 nPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheValT 218

|||||
203 TTTCGAGGTGGCCGGGAGAGTGGCCCAACCATGAAGAAGCTTTGTGA 252

218 hrLysValSerValGlyLupheValGlyGluGlyGlyLysSerLys 234

|||||
253 CCAAGGTTTCGTTGGGAGTTGTGGGGGAAGGTGAAGGGAAGCAAG 302

235 LysIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLysLys 251

|||||
303 AAGATTTCAAGAAATAATGCCCATAGCTGTTCTTGAGGAGCTGAAGAA 352

251 sLeuProProLeuProAlaValGluArgValLysProArgIleLysLysL 268

|||||
353 GTTACCGCCCTGCTGACAGTTGAACGAGTAAAGCCTAGAAATCAAAAAAGA 402

268 ySThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnGly 284

|||||
403 AAACAAAACCCATAGTCAAGCCACAGACAGCCCAAGATATGGCCAGGGG 452

285 IleAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlyLys 301

|||||
453 ATCAATCCGATTAGCCGACTGGCCAGATCCAGACAGGCAAGAAAAAGAGAA 502

301 sGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArg 318

|||||
503 GGAGCCAGAGTACAGCTCTCTCACAGAGCGAGGCTCCCGCGCGCAGGG 552

318 LuPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrGly 334

|||||
553 AGTTGTGATGCGAGGTGAAGTGTGAACACACACTGCAGAGGAACGGGC 602

335 ThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluI 351

|||||
603 ACCAACAGAAGGTGGCCAAAGCGCAATGACGCCGAGAGACATGCTGAGAT 652

351 eLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLysS 368
|||||
653 CTTGGTTCAAGTCCCGCAGCGCAGCCCAACCAACCGCACCTCAAGT 702
368 erGluGluLysThrProIleLysLysProGlyAspGlyArgLysValThr 384
|||||
703 CAGAGGAGAAGACACCATTAAGAAACACAGGGATGGAAGAAAGTAACC 752
385 PhePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAs 401
|||||
753 TTTTGTGAACCTGGCTCTGGGGATGAAAATGGAGACTAGTAATAAGAGCA 802
401 pGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleL 418
|||||
803 TGAGTTCAGGATGCCTTATCTAAGTCATCAGCAGCTGCCTGTGAATTC 852
418 euProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyHis 434
|||||
853 TTCCCATGGTCCGAGGTCCGCCAGGCTGTAGGAGTTAGTCAAGGACAT 902
435 HisThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrVa 451
|||||
903 CACACCAAGATTTCACAGGGCAGCTCCGAATCCTGCCAAGGCCACGGT 952
451 lThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrA 468
|||||
953 AACTGCCATGATAGCCCGAGAGTGTGTATGGGGGCACCTCGCCACAG 1002
468 lA.GluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisGl 484
|||||
1003 CCCGAGACCATTTTAAGAAATAACATCTCTTCAGGCCACGATACCCCATGG 1052
484 yProLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGlnG 501
|||||
1053 ACCTCTCAGCAGACCTCTTGAGCAACTGGACTATCTTNCAGAGTCCAGG 1102
501 lypheGlnValGluTyrLysasp 508
|| |||||
1103 GATTNCAGGTTGAATACCAAGAC 1125
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:X39992
seq_documentation_block:
ID X39992 standard; DNA; 773 BP.
XX
AC X39992;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

alignment_scores:
Quality: 1292.00 Length: 257
Ratio: 5.087 Gaps: 1
Percent Similarity: 98.833 Percent Identity: 97.665
alignment_block:
US-09-316-048-6 x X39992 ..
Align seg 1/1 to: X39992 from: 1 to: 773

253 ProProLeuProAlaValAluArgValLysProArgIleLysLysLysTh 269
|||||
3 CGGCCCTGCTGCAGTTGAACGAGTAAGCCTAGAAATCAAAAAAGAAAC 52
269 rLysProIleValLysProGlnThrSerProGluTyrGlyGlnGlyIleA 286
|||||
53 AAAACCCATAGTCAAGCCACAGACCAAGCCAGAAATATGGCCAGGGATCA 102
286 snProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGlu 302
|||||
103 ATCCGATTAGCCGACTGCGCCAGATCCAGCAGCAAAAGAGAGAGAG 152
303 ProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgGluPh 319
|||||
153 CCAGAGTACACGCTCCTCACAGACGAGGCCCTCCGCCGCCAGGAGTT 202
319 eValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyThra 336
|||||
203 TGTGATGCAGGTGAAGGTTGGAAACACACACTGCAGAGGAAGGCGCACCA 252
336 snLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeu 352
|||||
253 ACAAGAGGTGGCCCAAGCGCAATGCAGCCGAGAACATCTGGAGATCCTT 302
353 GlyPheLysValProGlnArgGlnProThrLysProAlaLeuLysSerGl 369
|||||
303 GGTTCAAAGTCCCGCAGCGCGCACCCACCAAAACCGCAGTCAAGTCAGA 352
369 uGluLysThrProIleLysLysProGlyAspGlyArgLysValThrPhep 386
|||||
353 GGAGAGACACCCATTAAGAAACACAGGGATGGAAGAAAGTAACCTTTT 402
386 heGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspGlu 402
|||||

403 TTGAACCTGGCTCTGGGGATGAAAATGGGACTAGTATAAAGAGCATGAG 452
403 PheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPr 419
453 TTCAGGATGCCTTATCTAAGTCATCAGCACTGCTGCTGGAATCTTCC 502
419 oMetValProGlnValAlaGlnAlaValGlyValSerGlnGlyHisHisT 436
503 CATGTGCCCGAGGTGCGCCAGCGTGTAGAGTTAGTCAAGGACATCACA 552
436 hrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThr 452
553 CCAAGATTTTACCAAGGGCAGCTCCGAATCTGCCAAGGCCAGGTAAC 602
453 AlaMetIleAlaArgGlnLeuLeuTyrGlyGlyThrSerProThrAla.G 469
603 GCCATGATAGCCCCGAGAGTGTGTATGGGGGCACCTCGCCCAACACCCCG 652
469 luThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisGlyPro 485
653 AGACCATTTTAAAGAAATACATCTCTTCAGGCCACGTACCCCATGGACCT 702
486 leuThrArgProSerGlnGlnLeuAspTyrLeuSerArgValGlnGlyPh 502
703 CTCACGAGACCCCTNTGAGCAACTGACATCTTTNCAGAGATCCAGGATTT 752
502 eGlnValGluTyrLysAsp 508
753 NCAGGTGGAATACCAAGAC 771

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X40000

seq_documentation_block:

ID X40000 standard; DNA; 844 BP.

AC X40000;

DT 02-JUL-1999 (first entry)

DE Prostate cancer associated gene.

KW Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer; ss.

OS Homo sapiens.

PN WO9904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98MO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

DR WPI; 1999-132448/11.

PT New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

alignment_scores:

Quality: 1257.50 Length: 281

Ratio: 4.800 Gaps: 2

Percent Similarity: 93.238 Percent Identity: 90.391

alignment_block:

US-09-316-048-6 x X40000 ..

Align seg 1/1 to: X40000 from: 1 to: 844

189 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 205
3 GTGTTTGAGATTGGCACTTAAACGGAACCTGCTGTGAATTTGAGGTGGC 52

205 aArgGluSerGlyProProHisMetLysAsnPheValThrLysValSerV 222
53 CCGGAGAGTGGCCCAACCCACATGAAGACTTTGTGACCAAGTTTCGG 102

222 aIGlyGluPheValIGlyGluGlyGlyLysSerLysLysIleSerLys 238
103 TTGGGAGTTTGTGGGGGAAGGTGAAGGGAAGCAAGAGATTTCAAAG 152

239 LysAsnAlaAlaIleAlaValLeuGluLeuLysLysLysLeuProLe 255
153 AAAAATGCCCGCATAGCTGTCTTGAGGAGCTGAAGAAGTTACCGCCCT 202

255 uProAlaValGluArgValLysProArgIleLysLysLysThrLysProI 272
203 GCCTGCAGTTGAACGAGTAAGCCTAGAAATCAAAAAGAAACAAACCCA 252

272 leValLysProGlnThrSerProGluTyrGlyGlnGlyIleAsnProIle 288
253 TAGTCAAGCCACAGACAAAGCCAGAAATATGCCAGGGATCAATCCGATT 302

289 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 305
303 AGCCGACTGGCCCAATCCAGCAGCAAGCAAAAAGAGAAGAGCCAGAGTA 352

305 rThrLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 322
353 CACGCTCTCACAGAGCAGGAGGCTCCCGCCGCGCAGGAGTTTGATGTC 402

322 InValLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLys 338
403 AGGTGAAGTTGAAACACACACTGCAGAAAGAACGGGCACCAACAAGAAG 452

339 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 355
453 GTGGCAAGCGCAATGCAGCCGAGAACATCTGGAGATCTTGTTTCAA 502

355 sValProGlnArgGlnProThrLysProAlaLeuLysSerGluGluLysT 372
503 AGTCCCGCAGCGCGCAGCCCAACCAACCGCACTCAAGTCAGAGAGAAGA 552

372 hrProIleLysLysProGlyAspGlyArgLysValThrPhePheGluPro 388
|||||
553 CAACCATAAAGAAACCAAGGGGATGGAGAGAAAGTAACCTTTTGAACCT 602
389 GlySerGlyAspGluAsnGlyThrSerAsnLysGluAspGluPheArgMe 405
::: |||||
603 GCTTGTGGGATGAAATGGGACTAGTAATAAGAGCATGAGTTGAGAT 652
405 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 422
|||||
653 GCCTTATCTAGTCATCAGCAGCTGCCCTGCGAATCTTCCATGGTGC 702
422 roGluValAlaGlnAlaValGly.ValSerGlnGlyHisThrLysAs 438
|| |||||
703 CCGANGTCGCCCAAGCTGAGGAAGTGTAGTCAAGACATNACACCAAGA 752
438 pPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetI 455
|||||
753 TTTTACCAGGACGCTT...CGAATCTTGCCANGCNCNGTACTGCCATGA 799
455 leAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAla 468
|||||
800 TAGCCCANAGTTGTGTATGTGGGCGANCCTTGGCCAGGCC 840

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X40002

seq_documentation_block:

ID X40002 standard; DNA; 946 BP.

XX X40002;

DT 02-JUL-1999 (first entry)

XX DE Prostate cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer; ss.

XX Homo sapiens.

XX WO9904265-A2.

XX 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

XX 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;

PI pFreundschuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

XX WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 629-630; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically

CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the

CC interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in

CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides

CC which are cancer associated antigen precursors expressed in human breast

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC lung cancer.

XX

SQ Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other;

alignment_scores:

Quality:	1255.50	Length:	312
Ratio:	4.516	Gaps:	3
Percent Similarity:	89.103	Percent Identity:	85.577

alignment_block:

US-09-316-048-6 x X40002 ..

Align seg 1/1 to: X40002 from: 1 to: 946

135 GlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 151
|||||
3 GGACAGCAATTATGCGCAAGGAAAGACAAAGACAGCGCTGCCAAACACGA 52
151 pAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluA 168
|||||
53 TGCTGTGCCAAAGCGTTGAGGATCTGCAGATGAGCCCTGCCAGAGA 102
168 rGLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSer 184
|||||
103 GGCTGAGGTGAATGGAAGAGAAATCCGAAGAGAAATACTCAATAATCT 152
185 GluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAs 201
|||||
153 GAAATAAGTCAAGTGTGTGAGATGCACCTTAACCGAAGCTTGCTGTGAA 202
201 npheGluValAlaArgGluSerGlyProProHisMetLysAsnPheValT 218
|||||
203 TTTCGAGGTGCCCGGGAGAGATGCCCAACCCACATGAGCAACTTTGTGA 252
218 hrLysValSerValGlyGluPheValGlyGluGlyLysSerLys 234
|||||
253 CCAAGGTTTCGGTTGGGAGTTGTGGGGGAAGGTGAAGGGAAGCAAG 302
235 LysIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLysL 251
|||||
303 AAGATTTCAAAGAAAAATGCCGCATAGCTGTTCTTGAGAGAGCTGAAGAA 352
251 sLeuProProLeuProAlaValAluArgValLysProArgIleLysL 268
|||||
353 GTTACCGCCCTGCTGCAGTTGACGAGTAAGCCTAGATCAAAAAAGA 402
268 ySThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnGly 284
|||||
403 AAACAAACCATAGTCAAGCCACAGACAAAGCCAGAAATATGCCAGGGG 452
285 IleAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlu 301
|||||
453 ATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGGCAAAAAAGAGAA 502
301 sGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgG 318
|||||
503 GGAGCCAGAGTACACGCTCTCACAGAGCGAGGCTTCCGCCCGCAGGG 552
318 lPheValMetGlnValLysValGlyAsnHisThr.AlaGluGlyThrG1 334
|||||
553 AGTTGTGATGCAAGTGAAGGTGGAACACACACTTGCAAGAGGAACGGG 602
334 yThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluI 351
|||||

XX 280269;
AC
XX 07-APR-2000 (first entry)
DT
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.
DE
XX Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
XX hyperplasia; ds.
XX Homo sapiens.
XX WO964576-A2.
XX 16-DEC-1999.
XX
XX 09-JUN-1999; 99WO-IB01062.
XX
XX 10-JUN-1998; 98US-0088801.
XX
XX (FARB) BAYER CORP.
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI; 2000-087220/07.
DR
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
XX Claim 15; Page 273; 469pp; English.
PS
XX
XX 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
CC
XX
XX Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;
SQ

alignment_scores:
Quality: 691.50 Length: 155
Ratio: 4.704 Gaps: 2
Percent Similarity: 94.839 Percent Identity: 90.323

alignment_block:
US-09-316-048-6 x 280269 ..

Align seg 1/1 to: 280269 from: 1 to: 727

116 TyrPheTyrProPheProValProProLeuLeuTyrGlnValGluLeuSe 132
111
3 TACTTTTACCCATTTCACGTTCCACCTTTACTTTATCAAGTGAACCTTTC 52
132 rValGlyGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAla 149
53 TGTGGAGAGACAGCAATTAAATGGCAAGAAAGACACAGAGGCTGCGA 102
149 ySHisAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeu 165
103 AACACGATGCTGCTGCCAAGCGTTGAGGATCCTGCAGAAATGAGCCCTG 152
166 ProGluArgLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAs 182

153 CCAGAGAGGCTGGAGGTGAATGAGAGAGAAATCCGAGAGAGAAATCTCAA 202
182 nLysSerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuP 199
203 TAAATCTGAATAATAGTCAAGTGTGTGAGATTGCACCTTAACGGAACCTGC 252
199 roValAsnPheGluValAlaArgGluSerGlyProProHisMetLysAsn 215
253 CTGTGAATTTTCGAGGTGGCCCGGAGAGAGTGCCACCCACATGAAGAAC 302
216 PheValThrLysValSerValGlyGluPheValGlyGluGlyGly 232
303 TTTGTGACCAAGGTTTCGGTTGGGAGTTTGTGGGGGAGGTGAAGGAA 352
232 sSerLysLysIleSerLysLysAsnAlaAlaIleAlaValLeuGluGlu 248
353 AAGCAAGAGATTTCAAAGAAATGCCCGCATAGCTGNTCTTGAGGAG 402
249 LeuLysLysLeuProProLeuProAlaVal...GluArgValLysProAr 264
403 CTGAAGAAAGTACCGNCCCTGCTTGNATTGGACCGCAAGCTTAAGCCTAN 452
264 gIleLysLysLys 268
453 AATCCAATGAA 465

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:217411

seq_documentation_block:

ID 217411 standard; cDNA; 769 BP.

AC 217411;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:4884.

KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

XX WO9938972-A2.

PD 05-AUG-1999.

PF 28-JAN-1999; 99WO-US01619.

PR 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

DR WPI; 1999-494092/41.

XX Novel human genes and their expression products which are
XX differentially expressed in different cell types

PT Claim 1; Page 2322; 2479pp; English.

PS The present invention describes a library of human polynucleotides
CC


```

167 TCCTTCTCGCCATGTCCAGAGACGAGCACC GCCCGCGCT 216
174 ggluSerGluGluAsnLeu..... 181
217 GGAGCGCGAGACAGTGGACCTTCAGTTTGGGGAAGATGATAACAGCTA 266
182 ....AsnlySerGluIleSerGlnValPheGluIleAlaLeuLys... 195
267 AGCCAGGGGAAAAACACCGATTTCAGGTATTACAGCAATACGGCATGAAGACC 316
196 ArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProH 212
317 AAGAACATCCAGTT...TATGAATGTGAAGATCTGATGTGCAAAATACA 363
212 smetLysAsnPheValThrLysValSerValGlyGluPheValGlyGlu 229
364 CGTGCCCACTTTCACCTTCAGAGTAAACCGTTGTCACATAACCTGCACAG 413
229 LysGluGlyLysSerLysLysIleSerLysLysAsnAlaAlaIleAlaVal 245
414 GTGAAGGTACAAGTAAAGACTGGCGAAACATAGAGCTGCAGAGGCTGCC 463
246 LeuGluGluLeuLysLysLeuProProLeuProAlaValGluArgVal 262
464 ATAAACATTTTGAAA.....GCCAA 483
262 sProArgIleLysLysThrLysProIleValLysProGlnThrSerP 279
484 TGCAGTATTTGCTTTCAGTTCCTGACCCCTTAATGCCCTGACCCCTTCCA 533
279 roGluTyrglyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGln 295
534 AGCAACCAAGAACAACAGCTTAATCTATTGTTTCATTACAGGAATTGGCT 583
296 GlnAlaLysLysGluLysGluProGluTyrThrLeuLeuThrGluArg 312
584 ATTTCATCATGGCTGGAGACTTCCTGATATATACCTTCCACGAGGAGG 633
312 yLeuProArgArgArgGluPheValMetGlnValLysValGlyAsnHis 329
634 ACCTGCTCATAGAGAGAAATATACTACAATTGGCAGGCTAGAGTCATTTA 683
329 hralaGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaAla 345
684 TGGAAACTGGAAAGGGGGCATCAAAAAAGCAAGCCAAAGGAATGTGCT 733
346 GluAsnMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProTh 362
734 GAGAAATTCTT..... 745
362 rLysProAlaLeuLysSerGluGluLysThrProIleLysLysProGly 379
745 ..... 745
379 spGlyArgLysValThrPhePheGluProGlySerGlyAspGluAsnGly 395
746 .....GCCAAATTAGTAATATTCTCCA.....GAGAACAC 778
396 ThrSerAsnLysGluAspGluPheArgMetProTyrlLeuSerHisGln 412
779 ATTTCT..... 784
412 nLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaVal 429
785 .....TTAACAAATGTAGTAG 800
429 LysValSerGlnGly.....HisHisThrLysAspPheThrArgAla 442
801 GACATTTCTTAGGATGACTTGCATTCCTTGAGGAAT..... 838
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459

```

```

838 ..... 838
459 uLeuTyrglyGlyThrSerProThrAlaGlu..ThrIleLeuLysAsnA 475
839 .....TCCTCTGGTGAAAGATCACTTACTGAAAGAA 872
475 snIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 491
873 GCCTCCTTAGT.....ATTCCAAT.....ACAGATTACATCCAG 907
492 GlnLeuAspTyrlLeuSerArgValGlnGlyPheGlnValGluTyrlLys 508
908 CTGCTTAGTGAATTTGCCAAGAACAAAGTTTAAATATAACATATTGGA 957
508 pPheProLysAsnAsnLysAsn...GluPheValSerLeuIleAsnCys 524
958 TATAGTGAACGTAGCGCCCATGGACAATATCAATGTCTGCTGAACGTGT 1007
524 erSerGlnProProLeuIleSerHisGlyIleGlyLysAspValGluSer 540
1008 CCACCAAGCCCATCACAGTGTGTATGCTCCGGTATCTCTGTGCAAT 1057
541 CysHisAspMetAlaAlaLeuAsnIleLeuLysLeu..... 553
1058 GCACAAAGTGATGCAGCTCAACAATGCTTTGCAGTATTAAAGATAATAGC 1107
554 .....SerGluLeuAspGlnGlnSerThrGluMetP 564
1108 AGAAGAAGTAATCTGAGCAACTTAAAAAATCTTCAGTAGACACATA 1157
564 roArgThrGlyAsnGlyPro 570
1158 AAAAGTTCCCTCTGCGCCC 1177

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:Z52512
seq_documentation_block:
ID Z52512 standard; cDNA; 1396 BP.
XX
AC Z52512;
XX
DT 29-FEB-2000 (first entry)
XX
DE Human secreted protein clone yj10_1 nucleotide sequence SEQ ID NO:75.
XX
KW Human; secreted protein; immunostimulatory; haemostatic; cytokine;
KW proliferative; differentiative; chemotactic; chemokinetic; vaccine;
KW thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9958642-A2.
XX
PD 18-NOV-1999.
XX
PF 14-MAY-1999; 99WO-0510843.
XX
PR 14-MAY-1998; 98US-0085472.
PR 17-AUG-1998; 98US-0096824.
PR 11-SEP-1998; 98US-0099843.
PR 11-SEP-1998; 98US-0099950.
PR 15-SEP-1998; 98US-0100424.
PR 29-SEP-1998; 98US-0102329.
PR 09-OCT-1998; 98US-0103615.
PR 11-DEC-1998; 98US-0111799.
PR 14-DEC-1998; 98US-0112159.
PR 31-DEC-1998; 98US-0114415.
PR 10-FEB-1999; 99US-0248059.
PR 06-APR-1999; 99US-0287150.
PR 13-MAY-1999; 99US-0311021.
XX
PA (GEMV ) GENETICS INST INC.

```


OM of: US-09-316-048-11 to: GenEmbl:* out_format : pfs
Date: Apr 5, 2001 10:36 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet.p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09316048/runat_01042001_170039_7742/app-query.fasta_1.1293
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdl -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09316048_@CGN1_1_9054 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-316-048-11
Query length: 487
Database: GenEmbl:*
Database sequences: 1118133
Database length: -1736092196
Search time (sec): 4765.960000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_ro:AF061942	+ 2550.00	2613.00	2.2e-137	2860	AF061942 Mus musculus RNA-bind
gb_ro:AF290989	+ 2512.00	2573.52	3.4e-135	3041	AF290989 Rattus norvegicus sta
gb_ro:AF227200	+ 2499.00	2560.13	1.9e-134	3059	AF227200 Rattus norvegicus sta
gb_pr3:AF061940	+ 2332.00	2388.61	6.9e-125	3142	AF061940 Homo sapiens staufer
gb_pr3:AF061938	+ 2332.00	2388.42	7.0e-125	3217	AF061938 Homo sapiens staufer
gb_pr3:AF061941	+ 2332.00	2388.09	7.3e-125	3349	AF061941 Homo sapiens staufer
gb_pr3:AF061939	+ 2332.00	2387.71	7.7e-125	3506	AF061939 Homo sapiens staufer
gb_pr4:AK025519	+ 2275.50	2330.52	1.2e-121	3194	AK025519 Homo sapiens cDNA: FI
gb_pr6:HSAL32258	+ 2274.50	2329.84	1.3e-121	3066	AJ132258 Homo sapiens mRNA for
gb_hg14:AC063967	- 1061.00	1049.65	2.6e-50	225016	AC063967 Mus musculus clone
gb_pr6:HSAS9062	+ 927.00	945.34	1.7e-44	4058	Y19062 Homo sapiens mRNA for 3
gb_pr4:AK002152	+ 914.00	934.59	6.7e-44	2968	AK002152 Homo sapiens cDNA FL
gb_pr6:HSB140E4	+ 705.50	687.65	3.8e-30	163253	AL121877 Human DNA sequence
gb_pr4:AL133174	- 701.00	686.01	4.7e-30	113836	AL133174 Human DNA sequence
gb_in3:DROSTAUFE	+ 634.50	643.21	1.1e-27	5237	M69111 D.melanogaster mRNA, co
gb_ro:RNO010200	+ 608.00	638.12	2.2e-27	360	AJ010200 Rattus norvegicus mRNA
gb_in2:AF225925	+ 601.50	605.86	1.4e-25	8002	AF225925 Musca domestica stau
gb_in1:AC004336	+ 599.50	587.01	1.5e-24	61267	AC004336 Drosophila melanoga
gb_hg6:AC017563	- 599.50	581.42	3.1e-24	120625	AC017563 Drosophila melanoga
gb_in1:AE003800	- 559.50	574.85	7.3e-24	267271	AE003800 Drosophila melanoga
gb_pr4:AK001576	+ 556.00	566.82	2.0e-23	3173	AK001576 Homo sapiens cDNA FL
gb_in2:AF225924	+ 544.00	546.47	2.8e-22	8409	AF225924 Drosophila virilis st
gb_pr4:AK023314	+ 453.00	464.95	9.7e-18	2008	AK023314 Homo sapiens cDNA FL
gb_hg14:AC055879	- 350.50	323.63	7.2e-10	160979	AC055879 Homo sapiens chromo
gb_hg3:AC011492	- 271.00	242.25	2.5e-05	157803	AC011492 Homo sapiens chromo
gb_hg15:AC068845	- 260.00	229.24	0.0001	194407	AC068845 Homo sapiens chromo
gb_hg14:AC055879	+ 258.00	228.75	0.0001	160979	AC055879 Homo sapiens chromo
gb_in2:CELE55A4	- 247.50	229.05	0.0001	42099	U67949 Caenorhabditis elegans
gb_pr8:HUMTRBP	+ 232.00	241.43	2.7e-05	1368	M60801 Human TAR RNA binding p
gb_hg12:AC027018	- 231.00	200.53	0.0052	171591	AC027018 Homo sapiens chromo
gb_hg22:AL390198	- 230.00	194.03	0.0119	332811	AL390198 Homo sapiens chromo
gb_pr7:HSU08998	+ 225.50	234.19	6.9e-05	1465	U08998 Homo sapiens TAR RNA bi
gb_pati:AR028451	+ 219.50	227.27	0.0002	1608	AR028451 Sequence 2 from paten
gb_pr3:AF072860	+ 219.00	226.11	0.0002	1740	AF072860 Homo sapiens protein
gb_pr6:HS265J14	- 217.50	191.95	0.0155	90547	Z84477 Human DNA sequence frc
gb_ro:MMU79962	+ 215.50	223.63	0.0003	1520	U79962 Mus musculus protamine-
gb_pr3:AF083033	+ 212.50	221.71	0.0003	1322	AF083033 Homo sapiens PKR-assc
gb_ro:AF083032	+ 209.50	217.14	0.0006	1584	AF083032 Mus musculus PKR-assc
gb_pr7:HSU10324	+ 206.00	206.99	0.0023	3505	U10324 Human nuclear factor NF
gb_ov:XLNRNABIPR	+ 199.00	207.64	0.0021	1358	M96370 Xenopus laevis RNA bindi

gb_pr4:AF167570	+ 198.50	199.94	0.0056	3242	AF167570 Homo sapiens doubl
gb_pr6:HS271747	+ 194.00	191.98	0.0155	4866	AJ271747 Homo sapiens parti
gb_pr7:HST000002	+ 193.50	203.25	0.0036	1166	AL078636 Homo sapiens mRNA
gb_pr4:AF167569	+ 193.00	193.85	0.0122	3426	AF167569 Homo sapiens doubl
gb_ro:AF098967	+ 193.00	193.85	0.0122	3426	AF098967 Mus musculus inter

seq_name: gb_ro:AF061942
seq_documentation_block:
LOCUS AF061942 2860 bp mRNA ROD 04-MAR-1999
DEFINITION Mus musculus RNA-binding protein staufer mRNA, complete cds.
ACCESSION AF061942
VERSION AF061942.1 GI:4335944
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2860)
Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.
Mammalian staufer is a double-stranded-RNA- and tubulin-binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell Biol. 19 (3), 2220-2230 (1999)
JOURNAL 99147057
MEDLINE 2 (bases 1 to 2860)
AUTHORS Duchaine,T., Luo,M. and DesGroselliers,L.
TITLES Direct Submission
JOURNAL Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES
source
1..2860
/organism="Mus musculus"
/db_xref="taxon:10090"
326..1789
/codon_start=1
/product="RNA-binding protein staufer"
/protein_id="A017529.1"
/db_xref="GI:4335945"
/translation="MYKVPDPHRSRMOSTYSYGMRGAYPRRYFPYPPVPLLQVELS
VGGOQFNGKGMPPVKHDAARALRTLOSEPLPERLEVNGRAEENLNKSEISQV
ETALRNLVNFVEVARESGPRPMKNFTVRVSGEVEGEGSKISKNAARAVLEQ
LRLPLPAVERVKKPRIKKSQPTCKTAPDYCGMNPISRLAIOQAKKEPEYML
TERGLPRREFVMQVKGHTHEGVGTNKVAKRNAEMLIEIGFKVPOAOPAKPAL
KSEBKTPTKPGDGKRVTFEPPSPGDENGTSNKDEEFMPYLSHQQLPAGILPMVPEV
AQAAGVSQGHHTKDTFTRAPNPATKATVTAMTARELLYGTSPTAETIKSNISSGHVP
HGPRTRPSEQLYLSRAQGFQVEYKDFPKNNKNECVSLINCSSQPLVSHGIGKDVES
CHDMAALNITKLSELDOQSTEMPTGNGPVSAACGR"

BASE COUNT	747 a	750 c	713 g	650 t
ORIGIN				
alignment_scores:				
Quality:	2550.00		Length:	487
Ratio:	5.236		Gaps:	0
Percent Similarity:	100.000		Percent Identity:	100.000

alignment_block:
US-09-316-048-11 x AF061942 ..
Align seg 1/1 to: AF061942 from: 1 to: 2860

1 MetTyrLysProValAspProHisSerArgMetGlnSerThyTyrSerTy 17
|||||
326 ATGTATAGGCCCGTGACCCCTCAGTGCAGTCCACCTACAGCTA 375
17 rGLyMetArgGLyGLyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
376 TGGCATGCGTGGAGGTGCTATACCCCGACATACCTTTACCCATTCCAG 425
34 alProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50
|||||
426 TCCACCTTACTCTACCAAGTGTGAGCTCTCCGTGGCGGACAGCAGTTT 475


```

/codon_start=1
/product="staufen isoform Stau-I6"
/protein_id="AAF98177.1"
/db_xref="GI:9755163"
/translation="MYKPVDPHSMOSTYSYGMRGAYPPRYFYPPVPLLQVELS
VGGQFNGKGMRRPVKHDATAARLRIOSEPLPERLEVNGRESEENLNKSEISOV
EIALKRNLPVNEVARESGPRPHMKNFVTVSVGEFVGESEKSKISKNAARAVLEQ
LRRLPLPAVERVKPRIKKSQPTCKLOTPADYGGMNPISRLAOIQAKKEPEYM
LITERGLPRRREVMQVKGHTAEGAGTNKKVAKRNAEENMLEIFGKVPQAPAKP
ALKSEKTPVKKPGDGRKVTFFEPSPGDENGTSNKEDEFMRPIYSHQQLPAGILPMVP
EVAQAVGVSOGHHTKDFARAAPNPAPAKATVTAMIARELLYGSTPDAETILKSNISSGH
VPHGPRTRPSEQIYLSRAQGFQVEYKDFPKNNKNECVSLINCSQPLVSHGIGKDV
ESCHDMAALNLIKLLSELDOQSTEMPRTCNGPVSACGRC"
polya_signal
3025..3030
/gene="Stau"
BASE COUNT      814 a      771 c      747 g      707 t      2 others
ORIGIN
```

alignment_scores:

```

Quality: 2512.00      Length: 489
Ratio: 5.169      Gaps: 1
Percent Similarity: 99.387      Percent Identity: 98.364
```

alignment_block:
US-09-316-048-11 x AF290989 ..

Align seg 1/1 to: AF290989 from: 1 to: 3041

```

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSery 17
|||||
227 ATGTATAGCCCGTGAGCCCGCCACTCTCGATGCAGTCCACTACAGCTA 276
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
277 TGGCATGCGTGAGAGTGCCCTATCCCGCCAGATACTTTACCAATTTCCAG 326
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50
|||||
327 TTCCACCGTTACTCTACCAAGTTGAGCTCTCCGTGGAGAGACAGCAGTTT 376
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
377 AATGGGAAGGGAAGATGAGACCACCTGTGAACATGATGCCACTGCCCG 426
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
|||||
427 AGCCCTGAGACTCTGCAGAGTGAGCCCTTACCAGAGAGGCTAGAGTTAA 476
84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
|||||
477 ATGGAAGAGATCAGAAGAAGAAACCTTAATAATCAGAAATAAGCCAA 526
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
527 GTGTTTGAGATTGCGCTGAAGCGGAATTGCGCTGTGAATTTTGAGGTGCC 576
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValServ 134
|||||
577 CCGGAGAGTGGCCCGCCACACATGAAGAACTTTGTGACCAGGGTTTCAG 626
134 aGlyGluPheValGlyGluGlyLysSerLysLysIleSerLys 150
|||||
627 TTGGGAGTTTGTAGGGGAAGGAGGAAAAAGCAAGACATCTCCAAG 676
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProLe 167
|||||
677 AAGAATGCGGCAAGGCGCGTGTGAGACAGCTAAGAGAGGCTGCCACCTCT 726
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnProt 184
|||||
727 ACCTGCTGTGAGCGAGTGAGGCCAGATCAAGAGAAAGTCAAGCCCA 776
```

```

184 hrCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||||
777 CCTGCAAGCTACAGACTGCCCCCGACTATGCCCCAAGGATGAATCTTATT 826
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215
|||||
827 AGCAGACTGGCAGATCCAGACAGGCAAAAAAGAGAAAGAGCCGAGCTA 876
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|||||
877 CATGCTCCTTACAGAACGAGGCTCTCCAGTCCGACGAGGAGTTTGATGC 926
232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
927 AGGTGAAGGTTGGGCATTCACACCCGACAGACGAGGATACCAATAGAAG 976
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 265
|||||
977 GTGCCCAAGCGCAATGCTGCTGAGAACATGCTGAGATCCTGGGGTTCAA 1026
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
|||||
1027 AGTCCCCCAGCGCAGCCTGCGCAAGCCAGCACTCAAAATCAGAAAGAGA 1076
282 hrProValLysLysProGlyAspGlyArgLysValThrPheGluPro 298
|||||
1077 CTCACAGTAAGAAGAACAGAGATGGAAGAAAGTAACATTTTGAACCT 1126
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
|||||
1127 AGCCCTGGGATGAATAATGGAACCAAGTAATAAGAGATGAGTTGAGAT 1176
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
|||||
1177 GCCCTATCTTAGCCATCAGCAGCTGCCAGCCGGGATTTCTCCCATGGTGC 1226
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
|||||
1227 CAGAGGTTGCCAGCGCTGCGGGGTAGTCAAGGACACACATACCAAGAT 1276
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetI 365
|||||
1277 TTCCGCAGGCGAGCTCCAAATCCTGCCAAGGCCACGCTAACCTGCCATGAT 1326
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382
|||||
1327 AGCCCGAGAAATGTTGTACGGGGGCACCTGCCCCACAGCCGAGACCATTT 1376
382 eLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
|||||
1377 TAAAGATTAACATCTCTTACAGACATGTACCCCATGAGACCTCGCACAAGA 1426
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGl 415
|||||
1427 CCCTCTGAGCAACGTACTACCTTTCCAGAGCCCGAGGATTCAGGTTGA 1476
415 uTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleA 432
|||||
1477 ATACAAGACTTCCCAAGAACAACAAGACGAGTGTGTATCTCTCATCA 1526
432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
|||||
1527 ACTGCTCTTACAGCCACCTCTCTGTCAGCCATGGCATCGGCAAGCATGTG 1576
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465
|||||
1577 GAGTCTCTGATGATATGCTGACATGACATCTTAAAGCTGCTGCTGA 1626
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
|||||
1627 GTTGACCAACAAGACACAGAGATGCCAAGGACAGAAACGACACAGTTT 1676
482 erAlaCysGlyArgCys 487
```

|||||
1677 CAGCGTGGCGGAGGTGC 1693

seq_name: gb_ro:AF227200

seq_documentation_block:

LOCUS AF227200 3059 bp mRNA ROD 09-AUG-2000
DEFINITION Rattus norvegicus staufen isoform Stau+I6 (Stau) mRNA, complete
cds, alternatively spliced.

ACCESSION AF227200
VERSION AF227200.1 GI:9754865

KEYWORDS
SOURCE Norway rat.

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 3059)

AUTHORS Monshausen,M., Putz,U., Rehbein,M., Schweizer,M.,
DesGroselliers,L., Kuhl,D., Richter,D. and Kindler,S.

TITLE Two rat brain staufen isoforms differentially bind RNA
Unpublished

REFERENCE 2 (bases 1 to 3059)
AUTHORS Monshausen,M. and Kindler,S.

TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) University of Hamburg, Institute for Cell
Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg
20246, Germany

FEATURES
source location/Qualifiers

1. .3059
/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/tissue_type="brain"

/dev_stage="adult"

1. .3059
/gene="Stau"

227. .1714
/gene="Stau"

/note="RNA-binding protein; similar to Drosophila staufen;
contains four RNA-binding domains; rStau+I6"

/codon_start=1
/product="staufen isoform Stau+I6"

/protein_id="AAF98119.1"

/db_xref="GI:9754866"

/translation="MYKPVDPHSRMOSTYSYGMRGAYPPRYFPYPPVPLLYQVELS
VGGQFNGKGMKRPVKHDATARALRLTQSEPLPERLEVNGRESEENLNKSEISQVF
EIALKRLNLPVNFESFPLTQVARESGPPHMKNFVTRVSGVFGESEENLNKSEISQVF
RAVLEQLRLRPLPAVERVKPRIRKKSQPTCKLQTAQPDYQGMNPSRLAQIQAKKE
KEPEYMLTERGLPRREFVMQYVGHHTAEGAGTNKVKARNAENMLEILGFYVPO
AOPAKPALKSEKTPVKKPGDGKRVTFEFPSPDENGTSNKEDFRMPYLSHQQLPAG
ILPMVPEVAQAVGVSQGHHTKDFARAAPNPAKATVTAMARELLYGTSPTAETLIK
NISGHPHGRPTRPSEQLYLSRAQGFQVEYKDFPKNNKNECVSLINCSQPLVSH
GIGKDVESCHDMAALNLIKLSLELDQSTEMPTGNGPVSACGRC"

polyA_signal 3043. .3048
/gene="Stau"

BASE COUNT 817 a 778 c 749 g 713 t 2 others

ORIGIN

alignment_scores:

Quality: 2499.00 Length: 495

Ratio: 5.142 Gaps: 2

Percent Similarity: 98.182 Percent Identity: 97.172

alignment_block:

US-09-316-048-11 x AF227200 ..

Align seg 1/1 to: AF227200 from: 1 to: 3059

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
227 ATGTATAAGCCCGTGGACCCCACTCTCGATGCGAGTCCACCTACAGCTA 276

17 rGlyMetLArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
277 TGGCATGCGTGAGAGGTGCCCTATCCCCCAGATACTTTTACCATTTCAC 326

34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
327 TTCACCGCTTACTCTACCAAGTTGAGCTCTCCGTGGAGAGACAGCAGTTT 376

51 AsnGlyLysGlyLysMetLArgProProValLysHisAspAlaProAla 67
377 AATGGAGAGGAAAGATGAGACACCTGTGAACATGATGCCACTGCCCG 426

67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
427 AGCCCTGAGGACTCTGCAGAGTGAGCCCTTACCAGAGAGGCTAGAGTTAA 476

84 sngLArgGluAlaGluGluGluAsnLeuAsnLysSerGlnLleSerGln 100
477 ATGGAAGAGATCAGAGAAGAAACCTTAATAATCAGAAATAAGCCAA 526

101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGlu.... 115
527 GTGTTGAGATTGCGCTGAGCGGAATTGCTGTAATTTTGAGTCTTT 576

116ValAlaArgGluSerGlyProProHisMetLysAsn 128
577 CCTCTGACACAGGTGGCCCGGAGAGAGTGCCCGCCGACATGAAGAAGT 626

128 heValThrArgValSerValGlyGluPheValGlyGluGlyLys 144
627 TTGTGACCAAGGTTTCAAGTTGGGAGTTTGTAGGGGAGAGAGGAGAA 676

145 SerLysLysLleSerLysLysAsnAlaAlaArgAlaValLeuGluGln 161
677 AGCAAGAAGATCTCCAAAGAAGATGCGCAAGGCGCTGAGAGAGCT 726

161 uArgArgLeuProProLeuProAlaValGluArgValLysProArgLle 178
727 AAGGAGCTGCCACCTCTACTGCTGTGAGCGAGTCAAGCCAGAAATCA 776

178 yslLysSerGlnProThrCysLys.....ThrAlaProAspTyrGly 192
777 AGAAGAAAGTCAAGCCACCTGCAAGCTACAGACTGCCCGGACTATGGC 826

193 GlnGlyMetAsnProLleSerArgLeuAlaGlnLleGlnGlnAlaLys 209
827 CAAGGATGAATCCTATTAGCAGACTGGCACAGATCCAGCAGCAAAAAA 876

209 sglLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuProArg 226
877 GGAGAAGGAGCCGGAGTACATGCTCTTACAGAACGAGGCTTCCACGTC 926

226 rArgGluPheValMetGlnValLysValGlyHisHisThrAlaGluGly 242
927 GCAGGAGTTTGTGATGACAGTGAAGTTGGGCAATCACACCGAGAAGGA 976

243 ValGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMet 259
977 GCAGGTACCAATAAGAGGTGGCCCAAGCGCAATGCTGTGAGACATGCT 1026

259 uGluLeuLeuGlyPheLysValProGlnAlaGlnProAlaLysProAla 276
1027 GGAGATCCTGGGGTTCAAGATCCCCAGGCGCAGCCTGCCAAGCCAGCAC 1076

276 euLysSerGluGluLysThrProValLysLysProGlyAspGlyArgLys 292
1077 TCAAAATCAGAAAGAGAGACTCAGTAAAGAAACAGAGATGGAAGGAAA 1126

293 ValThrPhePheGluProSerProGlyAspGluAsnGlyThrSerAsn 309
1127 GTAACTTTTGTGAACCTAGCCCTGGGATGAATAATGAACAGTAAATAA 1176

309 saspgluGlupheArqMetProTyrLeuSerHisGlnGlnLeuProAlaG 326
1177 AGAGGATGAGTTCAGGATGCCCTATCTTAGCCATCAGCAGCTGCCAGCG 1226
326 lyileleuPrometValProGluValAlaGlnAlaValGlyValSerGln 342
1227 GGATTTCTCCCATGGTGCAGAGGTTGCCAGGCTGTGGGGTTAGTCAA 1276
343 glyHisHisThrLysAspPheThrArgAlaAlaProAsnProAlaLysAl 359
1277 GGACACCATTACCAAGATTTCGCCAGGCGACGCTCCAATCCTGCCAAGCG 1326
359 aThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerp 376
1327 CACGGTAACGTGCCATGATGATGCCCGAGAATTGTTGTACGGGGCACCTCGC 1376
376 roThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyHisValPro 392
1377 CCACAGCCGAGACCATTTTAAAGATTAACATCTCTTCAGGACATGATACC 1426
393 HisGlyProArgThrArgProSerGlnGlnLeuTyrTyrLeuSerArgAl 409
1427 CATGGACCTCGCACAAAGACCTCTGAGCAACTGTACTACCTTCCAGAGC 1476
409 aGlnGlyPheGlnValGluTyrLysAspPheProLysAsnLysAsnG 426
1477 CCAGGGATTCCAGGTGTAATACAAAGACTTCCCGCAAGAACAACAAGAACG 1526
426 LucYsValSerLeuIleAsnGlySerSerSerGlnProProLeuValSerHis 442
1527 AGTGTGTATCTCTCATCACTGCTCTTCACAGCCACCTCTCGTCAGCCAT 1576
443 GlyIleGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnI 459
1577 GGCAATCGGCAAGGATGTGAGTCTGTCATGATATGGCTGCACCTGAACAT 1626
459 eLeuLysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgT 476
1627 CTTAAAGCTGCTGCTGAGTTGGACCAACAGACAGACAGAGATGCCAAGGA 1676
476 hrGlyAsnGlyProValSerAlaCysGlyArgCys 487
1677 CAGGAACGAGCACAGTTTCACGCGTGCAGAGGTGC 1711
seq_name: gb_pr3:AF061940
seq_documentation_block:
LOCUS AF061940 3142 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061940
VERSION AF061940.1 GI:4335950
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 3142)
Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroseillers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3142)
AUTHORS DesGroseillers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
Source location/Qualifiers
1..3142
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"

gene /map="20q13.1"
1..3142
/gene="STAU"
CDS 288..1778
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AADI7532.1"
/db_xref="GI:4335951"
/translation="MKLGKKPMYKPYDPSRMOSTYNYNMRGAYPPRYFPYPPVPL
LYQVELSVGGQGFNGKGTROAKHDAAKALRILQNEPLPERLEVNGRESEENLNK
SEISOVFELALKRNLPVNFVARESGPPHMKNFVTKVSGEFGEGSKSKSKNA
AIAVLEELKLPPLPAVERVKPRIKKTKPIVKPQTSPEYGGGINPISRLAQIQAKK
EKEPEYTLTERGLPRREFVMQVGNHTAEGTGNKRYAKRNAEENMLELGFVP
QROPTKPAKSEKTPIKKPGDKRVTFEFGSGDENGTSNKEDEFMRPYLSHQOLPA
GILPMVPEVAQAVGSOGHTTKDFTRAAPNPAKATVTAMTARELLYGTSPTAETLK
NMISGHVHPGLTRPSEQLDYSRYQGFQVEYKDFPKNKNFEVSLINCSSQPLIS
HGIGKDVESCHDMAALNLIKLSLSDQOSTEMPRTGNGPMSVCGRC"
BASE COUNT 862 a 778 c 733 g 769 t
ORIGIN
alignment_scores:
quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002
alignment_block:
US-09-316-048-11 x AF061940 ..
Align seg 1/1 to: AF061940 from: 1 to: 3142
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
117 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
359 CAACATGAGAGAGGTGCTTATCCCGGAGGTACTTTAACCATTTCCAG 408
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
409 TTCACACTTTACTTTATCAAGTGAACCTTCTGTGGGAGACAGCAATTT 458
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
459 AATGGCAAGGAAGAAAGACAAGACAGGCTGCGAAACACGATGCTGCCAA 508
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
509 AGCGTTGAGGATCCTGCAGAAATGAGCCCCCTGCCAGAGGCTGGAGGTGA 558
84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
559 ATGGAAGAGAAATCCGAAGAAGAAATCTCAATAATCTGAATAAAGTCAA 608
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
609 GTGTTTGAGATGTGCACCTTAACGGAACCTTGCCGTGTAATTCGAGGTGC 658
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValServ 134
659 CCGGAGAGATGGCCACCCACACATGAAGAACTTTGTGACCAAGGTTTCGG 708
134 aGlyGluPheValGlyGluGlyGlyLysSerLysLysIleSerLys 150
709 TTGGGGAGTTTGTGGGGAAGGTGAAGGGAAGCAAGCAAGATTTCAAAG 758
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProle 167
759 AAAAATGCCGCATAGCTGTTCTTGAGGAGACTGAAGAAGTTACCGCCCT 808

167 uProAlaValGluArgValLysProArgIleLysLysSerGlnProT 184
1809 GCCTCGAGTTGAACGAGTAAAGCCTAGATCAAAAAGAAAACAAACCA 858
184 hrcyslys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
859 TAGTCAAGCCACAGACAGACCAGCAATATGCCCAGGGGATCAATCCGATT 908
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215
909 AGCCGACTGGCCCAAGATCCAGCAGGCAAAAAGAGAGAGCCAGAGTA 958
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
959 CACGCTCCTCACAGAGCGAGGCTCCCGCCGCGCAGGAGTTTGTGATGC 1008
232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
1009 AGGTGAAGTTGGAACACACACTGCAGAGGAACGGGCACCAACAAGAAG 1058
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLys 265
1059 GTGGCCAAAGCCCAATGCAGCCGAGAACATGCTGGAGATCCTTGTTTCAA 1108
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
1109 AGTCCCGCAGCGGCGCCACCAACCCGCACTCAAGTCAGAGGAGAAGA 1158
282 hrProValLysLysProGlyAspGlyArgLysValThrPheGluPro 298
1159 CACCCATAAAGAAACCAAGGGGATGGAAGAAAGTAACCTTTTGTGAACCT 1208
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluLupheArgMe 315
1209 GGCTCTGGGATGAATAATGGGACTAGTAATAAGAGATGAGTTCAGGAT 1258
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPrometValP 332
1259 GCCTTATCTAAGTCATCAGCAGCTGCCCTGCGAATTCTTCCCATGCTGC 1308
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
1309 CCGAGGTCGCCCGCTGAGAGTTAGTCAAGGACATCACACCAAGAAT 1358
349 pheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetI 365
1359 TTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGGTAACCTCCATGAT 1408
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382
1409 AGCCCGAGAGTTGTTGATGGGGCACCTGCCCAACAGCCGAGACCATTT 1458
382 eulysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
1459 TAAAGATAATACATCTTTCAGGCCACGTAACCCATGGACCTCTCACGAGA 1508
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415
1509 CCCTCTGAGCAACTGGACTATCTTTCAGAGTCCAGGATTCACAGTTGA 1558
415 uTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleA 432
1559 ATACAAAGACTTCCCAAAACAAACAAAGCAATTTGTATCTTTATCA 1608
432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
1609 ATTGCTCTCTCAGCCACCTCTGATCAGCCATGGTATCGGCAAGATGTG 1658
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465
1659 GAGTCCTGCCATGATATGGCTGGCTGAACATCTTAAAGTTGCTGTCTGA 1708

465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
1709 GTTGACCAACAAGTACAGAGATGCCAAGAACAGAAACGAGACCAATGT 1758
482 eAlaCysGlyArgCys 487
1759 CTGTGTGTGGAGGTGC 1775
seq_name: gb_pr3:AF061938
seq_documentation_block:
LOCUS AF061938 3217 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061938
VERSION AF061938.1 GI:4335946
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 3217)
Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and Desgroselliers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3217)
Desgroselliers,L., Wickham,L. and Luo,M.
AUTHORS Direct Submission
TITLE Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
JOURNAL Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1. .3217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. .3217
/gene="STAU"
363. .1853
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AAD17530.1"
/db_xref="GI:4335947"
/translation="MKLGKKPMYKRPVDPYSRMQSTNYNNMRGAYPRRYEPPVPPL
LYOVELSVGGQFNGKGTROAAKHDAAKALRIQNEPLPERLEVNGRSEENLNK
SEISOVFEIALKRNLPVNFVARSEGRPHMKNFYTKVSVGEFVGEGRKSKSKNA
AIAVLEELKRLPLPAVERVKPRIKKTKPIVKPQTSPEYGGGINPISRLAQIOAKK
EKPEPYTLTERGLPRRREFVMQVKGNNHTAEGTGNKKVAKRNAENMLEILGFKVP
QROPTKPAKSEKTPYKKRGDGRKVTFFEPGSDENGTSNKEDEFKMPYLSHOOLPA
GILPMVPEVAQAVGSQGHHTKDFTRAAPNPAKATVTAMIARELLYGTSPTAETILK
NNTSSGHVPHGPLTRPSEQLDYLRSVQGFQVEYKDFPKNNKNEFVSLINCSQPLIS
HGIGKDVESCHDMAALNLIKLLSELDDQSTEMPRTGNGPMSVCGRC"

BASE COUNT 889 a 796 c 742 g 790 t
ORIGIN
alignment_scores:
Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002
alignment_block:
US-09-316-048-11 x AF061938 ..
Align seg 1/1 to: AF061938 from: 1 to: 3217
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||

384 ATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATACTA 433
17 rGlyMetArGlglyAlaTyrProProArGTYrPheTYrProPhePro 34
1334 GCGCTTACTAAGTATCAGACAGCAGCTGCGGAAACAGCATGCTGCTGCCAA 583
434 CAACATGAGAGAGGTGCTTATCCCGGAGGTACTTTTACCATTTCAG 483
34 aLProProLeuLeuTYrGlnValGluLeuSerValGlyGlnGlnPhe 50
484 TTCCACCTTTACTTTATCAAGTGAACCTTCTGTGGAGAGACAGCAATT 533
51 AsnGlyLysGlyLysMetArGProProValLysHisAspAlaProAlaAr 67
534 AATGGCAAGAGAAAGACACAGAGGCTGCGAAGACAGATGCTGCTGCCAA 583
67 gAlaLeuArGThrLeuGlnSerGluProLeuProGluArGLeuGluValA 84
584 AGCGTTGAGGATCCTGCAGATGAGCCCTGCGCAGAGAGGCTGGAGGTGA 633
84 snglyArGgluAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
634 ATGGAAGAGATCCGAAGAGAAAAATCTCAATAAATCTGAATAAGTCAA 683
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
684 GTGTTTGAGATTGCACCTAAACGGAACCTTGCTGTGAATTCGAGGTGGC 733
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
734 CCGGAGAGAGTGGCCACCCACATGAAGACTTTGTACCAAGGTTTCGG 783
134 aIlglyLupheValGlyGluGlyGlyLysSerLysIleSerLys 150
784 TTGGGGAGTTGTGGGGAGAGGTGAAGGGAAGCAAGAGATTTCAAAG 833
151 LysAsnAlaAlaArgAlaValLeuGlnLeuArgArgLeuProProLe 167
834 AAAAATGCCGCATAGCTGTCTTGAGAGCTGAAGAGTTTACCGCCCT 883
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnProT 184
884 GCTTGCACTTGAACGAGTAAAGCCTTAGAATCAAAAAGAAAAACCA 933
184 hrcysLys.....ThrAlaProAspTYrGlyGlnGlyMetAsnProIle 198
934 TAGTCAAGCCACAGACACAGCCAGCAATATATGGCCAGGATCAATCCGATT 983
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTY 215
984 AGCCGACTGGCCAGATCCAGACGAGCAAAAAAGAGAGAGCCAGAGTA 1033
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
1034 CACGCTCCTCACAGAGCGAGGCTCCCGCGCCGCGAGGAGTTGTGATGC 1083
232 lnValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
1084 AGTGGAAGGTTGGAACACCACTGCAGAAGGAACGGCACCAACAAGAG 1133
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLY 265
1134 GTGGCCAAGCGCAATGCAGCGGAGAACATGCTGGAGATCCTTGTTCAA 1183
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
1184 AGTCCCGCAGCGGACAGCCACCAACCCGCACTCAAGTCAGAGAGAGA 1233
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
1234 CACCATTAAGAAACACAGGGGATGGAAGAAAGTAACCTTTTGTGAACCT 1283
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgME 315
1284 GGTCTGTGGGATGAANAATGGAGCTAGTAATAAAGAGATGAGTTCAGGAT 1333

315 tProTYrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPrometValP 332
1334 GCGCTTACTAAGTATCAGACAGCAGCTGCTGGAATTCCTCCATGTGTC 1383
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
1384 CCGAGGTGCGCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGAT 1433
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIl 365
1434 TTTCACGAGGCGAGCTCCGAATCCTGCCAAGGCGACGGTAACTGCCATGAT 1483
365 eAlaArgGluLeuLeuTYrGlyGlyThrSerProThrAlaGluThrIleL 382
1484 AGCCCGAGAGTTGTTGTATGGGGCACTCGCCACAGCCGAGACCATTT 1533
382 euLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
1534 TAAAGATAACATCTCTTCAGGCGACGTACCCCATGGACCTTCACGAGA 1583
399 ProSerGluGlnLeuTYrTYrLeuSerArgAlaGlnGlyPheGlnValGl 415
1584 CCCTGTGAGCAACTGAGCTATCTTTCCAGAGTCCAGGATTCACAGTTGA 1633
415 uTYrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleA 432
1634 ATACAAGAACTTCCCAAAAACAACAAGAGAAATTTGTATCTTATCA 1683
432 snCYsSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
1684 ATGTCTCTCTCAGCCACCTGTGATCAGCATGTATCGCAAGATGTG 1733
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuSerGl 465
1734 GAGTCTTGCCATGATATGCGCTGCGCTGAACATCTTAAAGTTGCTGTGA 1783
465 uLeuAspGlnGlnSerThrGluMetProArgTYrGlyAsnGlyProValS 482
1784 GTTGACCAACAAGTACAGAGATGCCAAGACAGGAACGACCAATGT 1833
482 eAlaCysGlyArgCys 487
1834 CTGTGTGGGAGGTGC 1850
seq_name: gb_pr3:AF061941
seq_documentation_block: 3349 bp mRNA PRI 04-MAR-1999
LOCUS AF061941
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061941
VERSION AF061941.1 GI:4335952
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE 1 (bases 1 to 3349)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3349)
AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source location/Qualifiers
1..3349
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="20"
/map="20q13.1"
1..3349
/gene="STAU"
495..1985
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="stau protein"
/protein_id="A01753.1"
/db_xref="GI:4335953"
/translation="MKLGKKPMYKPVDPYSRMQSTNYNMRGAYPPRYEYEPVPPL
LYOVELSVGGOQENGKGTROAKHDAAKALILQNEPLPERLEVNGRSEENLNK
SEISQVEEIALKRNLPVNEVARESGPHMKNFYTKVSGFEVGESEKSKISKNA
AIAVLEELKLPPLPAVERKPRIKKTKPIVKPQTSPEYGOGINPISRLAQIOQAK
EKEPEYTLTERGLPRRREVMQVKNHTAEGTGNKKVAKRNAENMLEILGFVP
OROPTKPAKSEKTPIKKPGDGRKVTFFEPGSDENGTSNKEDEFMPYLSHQLPA
GILPMVPEVAQAVGVSQGHHTKDFTRAAPNPAKATVTAMIARELLYGTSPTAETLK
NNTSSGHVPHGPLTRPSEODLYLSRVQGFQVEKDEPKNNKNEFVSLINCSQPLIS
HGIGKDVESCHDMAALNLIKLSLSDQOSTEMPTGNGPMVCGRC"

BASE COUNT 913 a 838 c 777 g 821 t
ORIGIN

alignment_scores:
Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x AF061941 ..

Align seg 1/1 to: AF061941 from: 1 to: 3349

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
516 ATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCCTATACTA 565
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|::|
566 CAACATGAGAGAGAGTCTTATCCCCGAGTACTTTTACCCTATTTCCAG 615
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnPhe 50
|||||
616 TTCACACCTTACTTATCAAGTGAAGTCTTCTGTGGAGGACAGCAATT 665
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
|||||
666 AATGCCAAAGGAAGACAAGACAGAGCTGCGAAACACGATGCTGTCCTCAA 715
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
:|||||
716 AGCGTTGAGGATCCTGCAGATGAGACCCCTGCCAGAGAGCTGAGGTGA 765
84 sngLysArgGluAlaGluGluGluAsnLeuAsnLysSerGluLeuSerGln 100
|||||
766 ATGGAAGAGAAATCCGAGAAGAAATCTCAATAATCTGAATTAAGTCAA 815
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
816 GTGTTTGAGATTGCACCTTAACGGAACCTTCCCTGTGAATTTGAGGTGCG 865
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
|||||
866 CCGGAGAGTGGCCACCCACATGAAGAACTTTGTGACCAAGTTTCGG 915
134 aLgLyGluPheValGlyGlyGlyGlyLysSerLysLysLysLysLys 150
|||||
916 TTGGGAGATTGTGGGGGAAGGTGAAGGAAACAGCAAGAGATTCAAG 965
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProle 167
|||||

966 AAAAATGCCGCCATAGCTGTCTTGAGGAGCTGAAGAAGTTACCGCCCT 1015
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnProT 184
|||||
1016 GCCTGAGTTGAAGCAGTAAGCCTAGAAATCAAAAAAGAAACAAACCA 1065
184 hrcLysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||
1066 TAGTCAAGCCACAGACAAGCCAGAAATATGCGCCAGGGATCAATCGAAT 1115
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTy 215
|||||
1116 AGCCGACTGGCCAGATCCAGCAGGCAAAAAAGAGAAGAGCCAGAGTA 1165
215 rmetLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetG 232
|
1166 CACGCTCCTCACAGAGCGAGGCTCCCGCGCCAGGAGTTGTGATGC 1215
232 lValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
1216 AGGTGAAGTTGGAACACACACTGACAGAGAAGCGGACCAACAAGAAG 1265
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPhe 265
|||||
1266 GTGCCCAAGCGCAATGCACCGCAGAACATGCTGAGATCCTTGTTCAA 1315
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluLysST 282
|||||
1316 AGTCCCGCAGCGCGCAGCCACCAACCCGCACTCAAGTCAGAGAGAAGA 1365
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|||||
1366 CACCCATAAAGAAACACAGGGGATGGAAGAAAGTAACCTTTTGAACCT 1415
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluLysPheArg 315
:::
1416 GCCTCTGGGATGAATAATGGAGCTAGTAATAAAGAGATGAGTTCAGGAT 1465
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
|||||
1466 GCCTTATCTAAGTCAATCAGCAGCTGCTGGAATTCTTCCCATGCTGC 1515
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
|||||
1516 CCGAGGTGCGCCAGGCTGTAAGAGTTAGTCAAGAGATCACACCAAGAT 1565
349 pheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMet 365
|||||
1566 TTTACCAAGGCGAGTCCGAATCTCTGCCAAGGCCACGGTAACCTGCCAT 1615
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIle 382
|||||
1616 AGCCCGAGAGTTGTTGATGGGGCACCTCGCCACAGCCGAGACCATTT 1665
382 eulysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398
|||||
1666 TAAAGAATAACATCTTTCAGGCCACGTAACCCATGACCTCTCAGAGAG 1715
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415
|||||
1716 CCCTCTGAGCAACTGCACTATCTTCCAGAGTCCAGGATTCAGGTTGA 1765
415 uTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerIleAla 432
|||||
1766 ATACAAGACTTCCCAAAACACACAGACGAATTTGTATCTCTTATCA 1815
432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
|||||
1816 ATTGCTCTCTCAGCCACCTCTGATCAGCCATGCTATCGGCAAGATGTG 1865
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465
|||||
1866 GAGTCCTGCGCATGATATGGCTGCGCTGAACATCTTAAAGTTGCTGTCTGA 1915

465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
|||||
1916 GTTGACCAACAAGTACAGATGCCAAGAACGACGACCAATGT 1965
482 eAlaCysGlyArgCys 487
|||||
1966 CTGTGTGTGGAGGTGC 1982

seq_name: gb_pr3:AF061939

seq_documentation_block:

LOCUS AF061939 3506 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.

ACCESSION AF061939
VERSION AF061939.1 GI:4335948

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 3506)

AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum

JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)

MEDLINE 99147057

REFERENCE 2 (bases 1 to 3506)

AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada

FEATURES
source location/Qualifiers

1. .3506

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/map="20q13.1"

1. .3506

/gene="STAU"

409. .2142

/gene="STAU"

/function="RNA binding protein"

/note="alternatively spliced transcript"

/codon_start=1

/product="staufen protein"

/protein_id="AADI7531.1"

/db_xref="GI:4335949"

/translation="MSQVOVOVONSALSGSOLINKNOSLISQPLMSIPSTTSLS
ENAGRIQNSALPSASITSTSAASITPTVELNALCMKLGKPMYKVPDYSRMOST
YNYNMRGAYPPEFYEPFPVPPLLYQVELSVGQGFNGKGTROAKHDAAKALRLIL
QNEPLPERLEVNGRESEENLNKSEISQVEIALKRLNPVNEVARESPPHMKNFVT
KVSVEFVGESEGEKSKISKKNAAATAVLELKLPLPAVERVMPRIKKTKPIVKPQ
TSPEYGGINPISRLAOIQAKKEPEYTLTERGLPRREFVMQKVGNTAEGTG
TNKKVAKRNAENMLEILGFVPPQRPTRKALKSEKTPIKKPGDKKVTFFEPGSD
ENGTSNKDEFRMPYLSHQQLPAGILPMVPEVAQAVGVSQGHKDTTRAAPNPAKAT
VTAMIARELILYGTSPTAETILKNNISGHHVPHGPLTRPSEQLDYLRSVQGFQVEYKD
FPKNKNKNEFVSLINCCSOPPLISHGIGKDVESCHDMAALNIIKLSELDOOSTEMPR
TNGPMSVCGRC"

BASE COUNT 965 a 875 c 790 g 876 t

ORIGIN

alignment_scores:

Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:

US-09-316-048-11 x AF061939 ..

Align seg 1/1 to: AF061939 from: 1 to: 3506

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
673 ATGTATTAAGCCTGTGTGACCCCTTACTCTCGGATGCAGTCCACCTATACTA 722
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
723 CAACATGAGAGGAGGTGCTTATCCCGGAGGTACTTTTACCATTTCACAG 772
34 aLProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
|||||
773 TTCCACCTTACTTATCAAGTGAACCTTCTGTGGAGAGACAGCAATTT 822
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
823 AATGGCAAGAAAGAACAGACAGAGCTGCGAAACACAGATGCTGCTGCCAA 872
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
|||||
873 AGCGTTGAGGATCCTCGAGATGAGCCCTGCGACAGAGGCTGAGGTGA 922
84 sngLyArgGluAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
|||||
923 ATGGAAGAGAATCCGAGAAGAAATCTCAATTAATCTGAATAAAGTCAA 972
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
973 GTGTTGAGATTGCACTTAACGGAACCTTGCTGTGAATTTCAGAGGTGCC 1022
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValServ 134
|||||
1023 CCGGAGAGTGGCCACCCACACATGAAGAACTTGTGACCAAGTTTCGG 1072
134 aLgLyIuPheValGlyGluGlyGlyLysSerLysLysIleSerLys 150
|||||
1073 TTGGGAGTTTGTGGGGAAGTGAAGGAAAAAGCAAGATTTCAAAG 1122
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProle 167
|||||
1123 AAAAATGCCGCCATAGCTGTCTTGAAGAGCTGAAGAGTTTACCGCCCT 1172
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnProt 184
|||||
1173 GCCTGCAGTTGAACGAGTAAGCTTAGAATCAAAAAGAAAAACAACCA 1222
184 hrCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||||
1223 TAGTCAAGCCACAGACAAAGCCCAATATGGCCAGGGGATCAATCCGATT 1272
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTy 215
|||||
1273 AGCCGACTGGCCCCAGATCCAGACGCAAAAAAGAGAGAGACCAAGATA 1322
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|||||
1323 CACGCTCCTCACAGAGCGAGGCGCTCCCGCGCCGACGAGATTGTGATGC 1372
232 lNValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
1373 AGGTGAAGGTTGGAACACACACTGCAGAAGGAGGCGCACCAACAAGAAG 1422
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 265
|||||
1423 GTGGCCAGCGCAATGACGCCGAGAACATGCTGAGATCCTTGTTCAA 1472
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
|||||
1473 AGTCCCGCAGCGCAGCCACCAACCCGCACTCAAGTCAGAGAGAGAAGA 1522
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|||||
1523 CACCCATAAGAAACACAGGGGATGGAAGAAAGTAACCTTTTGAACCT 1572

299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
::: |||||||||||||||||||||||||||||||||||||||
1573 GGCTCTGGGATGAAAAATGGGACTAGTAATAAAGAGATGAGTTCAGGAT 1622
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPrometValP 332
|||||
1623 GCCTTATCTAAGTCATCAGCAGCTGCTGCTGGAATTCCTCCATGCTGC 1672
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
|||||
1673 CCGAGGTGCGCCAGCTGTAGAGATTAGTCAAGGACATCACACCAAGAT 1722
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIle 365
|||||
1723 TTACCAAGGCGCAGCTCCGAATCCTGCCAAGGCCACGGTAAGTCCATGAT 1772
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382
|||||
1773 AGCCCGAGAGTTGTTGATGGGGGCACCTCGCCACAGCCGAGACCATTT 1822
382 euLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
|||||
1823 TAAAGAATAACATCTCTTACAGGCACGTAACCCATGACCTCTCAGAGA 1872
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415
|||||
1873 CCCTCTGAGCAACTGACATATCTTCCAGAGTCCAGGATTCACAGTTGA 1922
415 uTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleA 432
|||||
1923 ATACAAGACTTCCCCAAAACAACAAGACGAATTTGTATCTCTTATCA 1972
432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
|||||
1973 ATTGCTCCTCTCAGCCACCTCTGATCAGCCATGTTATCGGCAAGATGTG 2022
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465
|||||
2023 GAGTCCTGCATGATATGGCTGCGCTGACATCTTAAAGTGGCTGTGA 2072
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProVal 482
|||||
2073 GTTGACCAACAAGTACAGAGATGCCAAGAACAGAAACGACCAATGT 2122
482 eAlaCysGlyArgCys 487
||:::|||||
2123 CTGTGTGTGGAGGTGC 2139
seq_name: gb_pr4:AK025519
seq_documentation_block:
LOCUS AK025519 3194 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens CDNA: FLJ21866 fis, clone HEP02379, highly similar to
AF061938 Homo sapiens staufen protein (STAU) mRNA.
ACCESSION AK025519
VERSION AK025519.1 GI:10438059
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 CDNA to mRNA, clone_lib:HEP
clone:HEP02379.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Kawabata,A., HikiJi,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 3194)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio

Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Departent of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1. .3194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP02379"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
misc_feature
1. .3194
/note="highly similar to AF061938 Homo sapiens staufen
protein (STAU) mRNA"
BASE COUNT 894 a 787 c 732 g 781 t
ORIGIN
alignment_scores:
Quality: 2275.50 Length: 489
Ratio: 4.801 Gaps: 2
Percent Similarity: 96.933 Percent Identity: 90.593
alignment_block:
US-09-316-048-11 x AK025519 ..
Align seg 1/1 to: AK025519 from: 1 to: 3194
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
353 ATGTATAAGCCTGTGACCCCTTACTCTCGATGCAGTCCACCTATTAAC 402
17 rGlyMetArgGlyAlaTyrProProArgTyrPheTyrProPhePro 34
||:::|||||
403 CAACATGAGAGAGGTGCTTATCCCCGAGGTACTTTTACCCTATTCAG 452
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnPhe 50
|||||
453 TTCCACCTTTACTTTATCAAGTGAAGTCTTCTGTGGAGAGACGCAATTT 502
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
503 AATGGCAAGGAAGAACACAGACAGGCTGCGAAACACGATGCTGCTCCAA 552
67 gaLalLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
:|||||
553 AGCGTTGAGGATCTGCAGAAATGAGCCCTGCCAGAGAGGCTGAGGTGA 602
84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluLeuSerGln 100
|||||
603 ATGGAAGAGAAATCCGAAGAAATCTCAATAATCTGAATTAAGTCAA 652
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
653 GTGTTGAGATTGCACCTTAACGGAACCTTGCCTGTGAATTCGAGGTGCG 702
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValServ 134
|||||
703 CCGGAGAGTGGCCACCCACATGAGAACTTTGTGACCAAGCTTTCGG 752
134 aLgLyGluPheValGlyGluGlyGlyLysSerLysLysIleSerLys 150
|||||
753 TTGGGAGATTGTGGGGAGAGGTGAAGGAAAGCAAGAAATTTCAAAG 802

151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgLeuProProLe 167
|||||
803 AAAAATGCCGCATAGCTGTCTTGAGGAGCTGAAGATTACCGCCCT 852
167 uProAlaValGluArgValLysProArgIleLysLysSerGlnPro. 183
|||||
853 GCCTGCAGTGAACGAGTAAGCCTAGAATCAAAAAGAAAACAAACCAT 902
184 .ThrCys..LysThrAlaProAspTyrGlyGlnGlyMetAsnProIleS 199
::: |||::: |||::: |||
903 AGTCAAGCCACAGACAGACGCCAGAAATATGCCAGGGGATCAATCCGATTA 952
199 erArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215
|||||
953 GCCGACTGGCCACAGATTCACAGCAGCAAAAAAGAGAGAGAGCCAGAGTAC 1002
216 MetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG1 232
|||||
1003 ACGCTCCTCACAGAGCGAGGCCCTCCCGCCGCGCAGGAGTTGTGATGCA 1052
232 nValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLysV 249
|||||
1053 GGTGAAGSTTGGAAACACACACTGCAGAAAGAACGGGCACCAACAAGAGG 1102
249 aAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLys 265
|||||
1103 TGGCCAAGCGCAATGACGCCAGAACATGCTGAGATCTTGGTTTCAA 1152
266 ValProGlnAlaGlnProAlaLysProAlaLeuLysSerGlnGluLysTh 282
|||||
1153 GTCCCGCAGGCGCAGCCACCAAAACCGCAGCTCAAGTCAGAGAGAAGAC 1202
282 rProValLysLysProGlyAspGlyArgLysValThrPheGluProS 299
|||||
1203 ACCCATTAAGAAACACAGG.GATGGAAGAAAAGTAACCTTTTGAACCTG 1251
299 erProGlyAspGluAsnGlyThrSerAsnLysAspGluLupheArgMet 315
::: |||
1252 GCTCTGGGATGAAAATGGGACTAGTAATAAAGAGATGATTCAGGATG 1301
316 ProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValPr 332
|||||
1302 CCTTATCTAAGTCATCAGCAGCTGCTGCTGGAATTTCTCCCATGTGCGC 1351
332 oGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 349
|||||
1352 CGAGGTGCCCGCAGCTGTAGAGTAGTCAAGGACATCACCAAGATT 1401
349 heThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIle 365
|||||
1402 TTACCAGGCGCAGCTCCGAATCTCGCAAGGCCACGGTAAGTCCATGATA 1451
366 AlaArgGluLeuLeuTyrGlyThrSerProThrAlaGluThrIleLe 382
|||||
1452 GCCCGAGAGTTGTTGATGGGGCACCCTGCCACAGCCGAGACCATTTT 1501
382 uLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArgP 399
|||||
1502 AAAGAATAACACCTCTCTCAGGCCACGTAACCCATGGACCTCTCACGAGAC 1551
399 roSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGlu 415
|||||
1552 CCTCTGAGCAACTGGACTATCTTCCAGAGTCCAGGATTCAGGTTGAA 1601
416 TyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleAs 432
|||||
1602 TACAAGAAGCTCCCAAAACAACAAGACGAATTTGTATCTCTTATCAA 1651
432 nCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspValG 449
|||||
1652 TTGCTCTCTCAGCCACCTCTGATCAGCCATGTGTATCGGCAAGATGTGG 1701

449 LuSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGlu 465
|||||
1702 AGTCCTGCCATGATATGGCTGCGCTGAACATCTTAAGTTGCTGTGAG 1751
466 LeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValSe 482
|||||
1752 TTGACCAACAAGATACAGAGATGCCAAGAACAAGAAACGACCAATGTC 1801
482 rAlaCysGlyArgCys 487
::: |||
1802 TGTGTGTGGGAGGTGC 1817

seq_name: gb_pr6: HSA132258

seq_documentation_block: LOCUS HSA132258 3066 bp mRNA PRI 14-MAY-1999
DEFINITION Homo sapiens mRNA for staufen protein, partial.
ACCESSION AJ132258
VERSION AJ132258.1 GI:4572587
KEYWORDS staufen gene; staufen protein.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3066)
AUTHORS Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
TITLE Interaction of influenza virus NS1 protein and the human homologue of staufen in vivo and in vitro
JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE 99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS Ortin,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology, Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN

FEATURES
source location/Qualifiers
1..3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
36..1526
/gene="staufen"
36..1526
/gene="staufen"
/function="dsRNA binding protein"
/codon_start=1
/product="staufen protein"
/protein_id="CAB40082.1"
/db_xref="GI:4572588"

CDS

/translation="MKLGKKPMYKPPVDPSRMOSTYNNMRCGAYPPRYFPVPPPL
LYQVELSVGGQGFNGKGTROAAKHDAAKALRILQNEPLPERLENGRESEENLNK
SEISQVEFIALKRLNPVNEVARESGPPHMKNFVTKSVGFEVGESEKSKISKNA
AIAVLEELKLPPLPAVERVVKPRIKKTKPIVKPQTSPEYGGGINPISRLAQIOAKK
EKEPEYTLTERGLPRRREFFVMQKVGNHTAEGTNNKVAKRNAENMLEILGFKVP
QRSHQTRTOVRGEDTHKETRGWKKSNLFLNALGMENGTSNKEDEFRRMPYLSHQOLPA
GILPMVPEVAQAVGVSGHHTKDFTRAAPNPAKATVTAMIARELYLGGTSPYAEIILK
NNISSGHVPHGPLTRPSEQLDYLSRVQGFQVEYKDFPKNNKNEFVSLINCSSQPLIS
HGIGKDVESCHDMAALNIILKLSLDDQSTEMPRRTGNGPMSVCGRC"

BASE COUNT 891 a 693 c 699 g 783 t
ORIGIN

alignment_scores:
Quality: 2274.50 Length: 492
Ratio: 4.819 Gaps: 5
Percent Similarity: 95.935 Percent Identity: 90.244

alignment_block:
us-09-316-048-11 x HSA132258 ..

Align seg 1/1 to: HSA132258 from: 1 to: 3066
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||

```
57 ATGTATAGCCTGTTGACCCCTTACTCTCGGATGCAGTCCACCTATACTA 106
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPheProy 34
107 CAACATGAGAGAGGTGCTTATCCCCGAGGTACTTTTACCATTCCAG 156
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
157 TTCCACCTTACTTATCAAGTGAAGTCTTCTGTGGAGAGACAGCAATT 206
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
207 AATGGCAAGGAAGAACAGACAGAGGCTGCCGAACACGATGCTGCTCCA 256
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
257 AGCGTTGAGGATCTTCAGAAATGAGCCCTGCCAGAGAGGCTGAGGTGA 306
84 sngLyArgGluAlaGluGluGluAsnLeuAsnLysSerGluLeuSerGln 100
307 ATGGAAGAGAATCCGAGAAAGAAATCTCAATAAATCTGAATTAAGTCAA 356
101 valPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluVal 117
357 GTGTTTGAGATTGCACTTAAACGGAACCTGCTGTGAATTCGAGGTGGC 406
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSer 134
407 CCGGGAGAGTGGCCCAACCCACATGAGAACTTGTGACCAAGGTTTCGG 456
134 aIGlyGluPheValGlyGluGlyLysSerLysLysIleSerLys 150
457 TTGGGGAGTTGTGGGGGAGGTGAAGGGAAGAACAGAGATTTCAAAG 506
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgLeuProProle 167
507 AAAAATGCCGCATAGCTGTCTTGAGGAGCTGAAGAGTTTACCCTCCCT 556
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnPro 184
557 GCCTGCAGTTGAACGAGTAAAGCCTAGAAATCAAAAGAAACAAACCCA 606
184 hrcLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 198
607 TAGTCAAGCCACAGACAAAGCCAGAAATATGGCCAGGGGATCAATCCG 656
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGlu 215
657 AGCCGACTGGCCAGATCCAGCAGGCAAAAGAGAGAGCCAGAGTA 706
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMet 232
707 CACGCTCCTCACAGAGCGAGGCTCCCGCCGAGGAGTTGTGATGC 756
232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLys 248
757 AGGTGAAGTTGGAACACACACTGCAGAGGAACGGGCACCAACAAGAAG 806
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPhe 265
807 GTGGCCAAAGCGCAATGCAGCCGAGAACATGCTGAGATCCTTGTTCAA 856
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluLys 281
857 AGTCCCGCAGCGCA...GCCACCAAAACCCGACTCAAGTCAGAGAGAAG 903
282 ThrProValLysLysProGlyAspGlyArgLysValThrPhePheGlu 298
904 ACACCCATTAAGAAACAGGCGGATGGAAGAAAGTAACCTTTTGAAC 953
298 roSerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPhe 314
954 CTGGCTCTGGGATGGAATAATGGACTAGTAATAAAGAGATGAGTTTCAG 1003
```

```
314 gMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMet 331
1004 GATGCCCTTATCTAAGTCATCAGCAGCTGCTGCTGGAATTCCTCCATGG 1053
331 aLProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLys 347
1054 TGCCCGAGGTGCGCCAGGCTGTAGAGATTAGTCAAGACATCACACCAA 1103
348 AspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAla 364
1104 GATTTTACCAGGCGAGCTCCGAATCCTGCCAAGCCACGTAACCTGCCAT 1153
364 tIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThr 381
1154 GATAGCCCGAGAGTGTGTATGGGGCAGCTCGCCACAGCCAGACCA 1203
381 lLeuLysSerAsnIleSerSerGlyHisValProHisGlyProArgThr 397
1204 TTTTAAGAATAACATCTCTTCAGGCCACGTACCCCATGAGACCTTCAC 1253
398 ArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGln 414
1254 AGACCCCTGAGACCACTGACTATCTTCCAGAGTCCAGGATTCAGGT 1303
414 lGluTyrLysAspPheProLysAsnAsnLysAsnGlyCysValSerLeu 431
1304 TGAATACAAAGACTTCCCAAAACAAACAGAAATTTGTATCTCTTA 1353
431 leasnCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAsp 447
1354 TCAATTGCTCCTGCACGCCCTGTGATGACCATGATCGGCAAGGAT 1403
448 ValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuSe 464
1404 GTGAGTCTCCTGCCATGATATGGCTGCGCTGAACATCTTAAAGTGTCT 1453
464 rGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyPro 481
1454 TGAGTTGACCAACAAGTACAGAGATGCCAAGAAGAAACGACCAA 1503
481 aISerAlaCysGlyArgCys 487
1504 TGTCTGTGTGTGGAGGTGC 1523

seq_name: gb_htg14:AC063967

seq_documentation_block:
LOCUS AC063967 225016 bp DNA 06-SEP-2000
DEFINITION Mus musculus clone RP23-41G23, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC063967 GI:9972323
VERSION AC063967.3 HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 225016)
McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Shekher,M., Spiegel,L.A., Toth,K. and Vill,M.D.
Mouse Genomic Sequence
JOURNAL Unpublished
TITLE 2 (bases 1 to 225016)
REFERENCE McCombie,W.R.
AUTHORS Direct Submission
JOURNAL Submitted (22-APR-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT On Sep 6, 2000 this sequence version replaced gi:8439872.
----- Genome Center
```


88193 GCACATGTGAACGTGCAGACAAAACACTCACATACAGAAGATAGAA 88144
186 186
88143 TCTTTATTCAGACATTGGCCCCCTTCCGCACTAGAAAGGACAGACTT 88094
186 186
88093 CAGCCAATCTCTCCTTTTTCATCTGGTGTCTCTGAAAATTACTTCCATC 88044
186 186
88043 ATCCCTGACAGTGCATGGGGAAGAGAGACCCCTAGTGTCTCAGAA 87994
186 186
87993 GCTTGACGTCCTCTCTCCCGGCGCTCACCTGGAGGTTTAAATTCATA 87944
186 186
87943 ACAGTGTTTGTTTGTCTACTGGGATTCGATGAGAGACTCAGGTGTGC 87894
186 186
87893 TGACCAAGCAGGCGCTCTTCCATAGGATGCATCTTCACACTGAGGCGTT 87844
186 186
87843 TACAAGGCTTTAAGTGTGTCTGGGTTTAAATGATTGAGAAAGCTAGGC 87794
186 186
87793 ATGGTCCCCCATGTAGATTTCAGTAGCGACAGGTCAAGCCCTGGTTCCAT 87744
186 186
87743 CCTCCCTTTTGTCAATCTTTTTTTTTTTTTTCTTGGAACTCACCTCT 87694
186 186
87693 GTAGACCAGGCTGGCCTCGAATTTCAGAAATCCGCCCTGCCCTCCACTCCCA 87644
186 186
87643 AGTTCTGGATTAAAGGCGTGCACCACCACTGCCCGCTCTTTTGTCA 87594
186 186
87593 TTCTCGCCAGTGTGTCTTACAGCTTTTACATGGCAGTTCACACTGTGC 87544
186 186
87543 TAATGTCCCAGTTACATGAGTACAGTTCGCCCATGCCGAGAGTGACAGA 87494
186 186
87493 GGTACAGGCAACACCAACCAAGAGACATCTCAGCATGATCTGAAC 87444
186 186
87443 ATGTTTAGTCTTAGAGAGGGTTTGGTCTGAGCGTTTACACCAGCACTA 87394
186 186
87393 GGTAGGTTTCTGTCTCAGTCCCGTGCAGCTGGACTGGCTGGTTCTTC 87344
186 186
87343 TCTGTGAGTAACACATATTGTCTCTGCACCTGCTGTTTTCAGCTACAG 87294
187 ThrAlaProAspTyrGlyGlnGlyMetAsnProIleSerArgLeuAlaG 203
|||||

87293 ACAGCCCCGGAATTATGCCCAGGAGTGAATCCTATTAGTAGACTTCACA 87244
203 nileGlnIleAlaLysLysGluLysGluProGluTyrMetLeuThrG 220
|||||
87243 GATCCAGCAGGCAGCAAAAAGAGAGAGAGCAGATACATGCTCTTACAG 87194
220 LuArgGlyLeuProArgArgArgGluPheValMet..... 231
|||||
87193 AACGAGTCTTCCAGCTCGCAGGAGTTGTGATGCAGGTGAGTTGGTGG 87144
231 231
87143 GGTGTAGGCAGCCCTTCCCTGACCATGTCTTAATCCGGCAGTTTCAG 87094
231 231
87093 TTAGCTGACTTGGTAGTGCCCTGCTTGCTCTCTGAGGCCAAGGCAA 87044
231 231
87043 GCGCAGGTGATCTCTATGAGTTCAAGACTCCACATTAGTGATCCCA 86994
231 231
86993 TCACAGAAACAGATCCACTCTGTAAGAGTGAGTTGACCAGAGTTGCC 86944
231 231
86943 TGAAGTGGCTGGGGACAGCTTTGATGAGTGTGCTGAATTTGTCCAGC 86894
231 231
86893 AGCCAGCCCTCAGAGAGAGTGTATGAGGAGAGACCCCACTGCACCTAG 86844
231 231
86843 ATATATACCTGTCTTAATAATTGATGATGTGTTGTGGACACATTTC 86794
231 231
86793 CATACATAGTTTAAAGTTTAAAGAACTCTTAAAGACTGTGAGAT 86744
231 231
86743 GGCTCAGTGTTAAGACACTGACTGCTCTCCGAAAGTCCTGAGTTCAA 86694
231 231
86693 ATCCAGCGACCATGTGGCTCACACCAACCCATATGAGATCTGATG 86644
231 231
86643 CCCTCTTCTGTGCATCTGAAGACAGCTACAGTACTTAATAATAAGTCT 86594
231 231
86593 TTAATAAAATAAATCTTTTAAAAATAGAGTTTCAATATCTTAACGT 86544
231 231
86543 GTGTGGCCATCACAGCAGTAGTAGTTCAGATAACGAACAATAAGCTT 86494
231 231
86493 TTTACCATCAAGTTTCAGCAGTATCCCTGGGGCTTTGGGTTAAGTCACAG 86444
232 GlnValLysValGlyHisIsthrAlaGlu 241
|||||
86443 GTCTTTGTTGTGTTTACACAGGTAAAGTTGGGCATCACACTGCAGAA 86394
242 GlyValGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMe 258
|||||
86393 GGAGTGGTACCAATAAGAAGGTGGCCAAGCGTAATGCTGCTGAGAACAT 86344

258 tLeuGluIleLeuGlyPheLysValProGlnAlaGlnProAlaLysProA 275
|||||
86343 GCTGAGATCCTGGGCTCAAGTAGTCCCGAGCGCAGCTGCCAAGCCAG 86294
275 lAlLeuLysSerGluGlu.Lys..... 281
|||||
86293 CACTCAATCAGAGAAGAGAGTAGTGTGGACTCTGTGCCAGGTGGCC 86244
281 281
86243 TAACTCCCTGCATGGGCAAGCCTGGGGCAAGGCTTCATCTCCCTGCAGT 86194
281 281
86193 AAGCGCCTGCCCGGGGCTCCTCCATTGAGATTAGATTGACAGAGG 86144
281 281
86143 GAGTTTCTGTCCAGCTCGTCACATCCTGGAGCCATCCATTCCGTTAAG 86094
281 281
86093 TCTCTCTTTAGTGTACACTATGGGCTAGATGTGCATCTCCGAGAGGTC 86044
281 281
86043 CTGCTCTCTACTTCTGATGGGTAGTTGGGGTTGGAGGAACTAAATGAC 85994
281 281
85993 TGGCACCTGTCTGCATGACCAGTTTGAATATGCCTCTGTTTCTCTGG 85944
281 281
85943 GACAAGCTAGCATGGTCAAAATAGCCAAATCTATAGATTGAGTTGAATA 85894
281 281
85893 TCACATACACCACCATTTGCTGGCTACAGCAGTTGTCAATTTGGGGATTG 85844
281 281
85843 AGCTTCTGAGCTCTGCCAGTTTCATTACAAGAAATAGCTCCCTCCACT 85794
281 281
85793 GCTGGAGCTTAGTACTTGGTACTCGCCTTGATTWGGCTTCCAGATTTC 85744
281 281
85743 CATTCAAAACAGTGTAGAGTCTATTAGTTCAGAGACTCCACATAGTGAGA 85694
281 281
85693 TCCCGTCACAGAAACAGTATCCGCTCTGTAAAGAGTGAGTTGGACCAGAGT 85644
281 281
85643 TTGCTGAGAGTGGCTGGGGGACAGCTTTGATGAAGTGTGCTGGAATCTGT 85594
281 281
85593 CTCTCTTGTCTGATAGCATTAGTACCTCTGTGTTCTTTGTGGATGTAC 85544
281 281
85543 TGTTTTGGATGTTCTAGATTATCTAAAGTGTTCGTGCTTAGCACAGTGC 85494
281 281
85493 AGACCTACTTTCCCAACAAGCTGAACACTTTTAAAGCCAGCATTTTCTCC 85444

281 281
85443 TGGCAGTGGTTTCTGTGTTCTTGCGACAGTGCCAGCTATGCATTCAAGAT 85394
281 281
85393 CTGCTCTGTGCAAAATGCTGCCATGGGGAGGCGGAGCGTTTGTCTATCCA 85344
282T 282
85343 CACAGACCCAGTGCTTGTCTATGCTCACCTGCCATTTTCTTCTTAAGA 85294
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|||||
85293 CTCCAGTAAGAAACCAGAGAGACGGAAGAAAGTAACGTTTTTTGAACCT 85244
299 SerProGlyAspGluAsnGlyThr..... 306
|||||
85243 AGCCCTGGGGATGAAATGGAACCTAGTAAGTGTGGCCCTCTGTCAATTG 85194
306 306
85193 AACCAATGTCTCTGTGCTTGTGTGGCTCTCGTAGATTTTTGCAAGCACA 85144
306 306
85143 GTATTCCTGCCCTTCCCTGTATGCTGGGATGTAAGGTTGGCTCTGTAGC 85094
307Ser..Asn 308
85093 TTCTTGGGTGGTGTATGGTACTTTGGTGACCTGTGTCCCTGCAGGTAAAC 85044
309 LysAspGluGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAl 325
|||||
85043 AAGACGAGAGGAGTTACAGATGCCTTATCTTAGCCATCAGCAGCTGCCAGC 84994
325 aglyIleLeuProMetValProGluValAlaGlnAlaValGlyValSerG 342
|||||
84993 TGGAACTTCCCATGATGGTCCGGAAGTTGCCAGGCTGTGGGGTTAGTC 84944
342 lnglyHisHisThrLysAspPheThrArgAlaAlaProAsnProAlaLys 358
|||||
84943 AAGACACACACCAAAAGATTTCACACAGGCGAGCTCCAATCTCGCCACAG 84894
359 AlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrse 375
|||||
84893 GCAACGGTAACTGCCATGATAGCCCCAGAGTGTGTACGGGGGCACCTC 84844
375 rProThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyHisValP 392
|||||
84843 GCCCAGAGCCGAGACCATTTTAAAGAGTAACATCTCTTCAGGCCACAGTAC 84794
392 roHisGlyProArgThrArgProSerGluGlnLeuTyrTyrLeuSerArg 408
|||||
84793 CCCATGGACCTCGCACTAGACCTCTGAGCAACTGTACTACTTTCACAGA 84744
409 AlaGlnGlyPhe.Gln..... 413
|||||
84743 GCCCAGGATTCACAGTAACCGTGTGTCATCTCTGGGCCCTTACGTTCTT 84694
413 413
84693 CCTGCACAGATTGCTTGTGTAGTGTGCAAGACCCTGGCTCTGTCTCT 84644
413 413
84643 CAGCACTACAGAAACAGAAAGGCAAGATTATATGCTTTATTAAAGATT 84594
413 413
84593 TGAATGTACAGTGAAGTTCTGAAGGAAGGAAGATGTCTGGGATGAGAATG 84544
413 413

```

84543 GCTTATATAAAGTAGCTATTCTTAAGCCCAACAACCTTGAGTATGATTA 84494
413 ..... 413
84493 CTGGTGGAAAGAACCAACTCCAGCAAGTTGTCTCAGTGTGCACACCCCTG 84444
413 ..... 413
84443 CCCCAGAGGTTACTGTGCGCCAGGGCTGAAGGAGAGAAGACACTTTT 84394
413 ..... 413
84393 ACTTAGGAAGCAGCCACTACATGAGCATGGAGAGAAGGCTTGAGGGTTC 84344
414 .....ValGluTy 416
84343 TGGGCTGGGCCACATGCAAGAGGTGTGTTTTTCTTAGGTTGAATA 84294
416 rlySaspheProLysAsnLysAsnGluCysValSerLeuIleAsn 433
84293 CAAAGATTTTCCCAAGACAACAAGAAGAGAGTGTATCTCATCAACT 84244
433 ySserSerGlnProLeuValSerHisGlyIleGlyLysAspValGlu 449
84243 GCTCCTCACAGCCGCTCTCGTCAGTCATGCCATCGGCAAGGATGTGGAG 84194
450 SerCysHisaspMet..... 454
84193 TCCTGTCAATGATATGATATGTGTGGCCAGAGGTTGTGAGAAGAGCTGGC 84144
454 ..... 454
84143 ACATTCCCTTGTTCAGTCAGTGTGTAGTGGCCCTCAGAAAGAGGACATG 84094
454 ..... 454
84093 GTTCTCCTGTAGCCTGTGGACATGTCTGATTGATTTCATTCTTCTCT 84044
454 ..... 454
84043 TCCTTGCTTCTTCTCTTCTTCCCTCACTTCTTGCGAGCTGTAGAAAGCCCC 83994
454 ..... 454
83993 TTTGGGAGTCTGTTCAAAGAAGGAATCAGCACACTTGAGCTTWTGGTGT 83944
454 ..... 454
83943 TCACTCGATGGACCAATAATGCACCCTTAAGCGGGGCACCCCTGTGATTG 83894
454 ..... 454
83893 AAGGGAACATGGGAGTTTGATGTTGAGCAGGATTCAAAACATTTCATTGG 83844
454 ..... 454
83843 GAAGTTTAAAGAGCTGGACTAGTCTGTTCCTGTGCTTGTGATTATTCTCT 83794
455 .....Ala..AlaLeuAsnIleLeuLysLeuLeuSerGluL 466
83793 TCCTTACACCTTTCCTAGGCTGCACCTGAACATTTTAAAGCTGTCTGAGT 83744
466 euAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValSer 482
83743 TGGACCAACAGAGACAGAGATGCCAAAGAACAGAAATGACACAGTTTCA 83694
483 Ala 483
83693 GCG 83691
seq_name: gb_pr6:HSA9062

```

```
seq_documentation_block:
LOCUS       HSA9062          4058 bp      mRNA                      PRI           14-JAN-2000
DEFINITION  Homo sapiens mRNA for 39k3 protein.
ACCESSION   Y19062
VERSION     Y19062.1 GI:6634456
KEYWORDS    39k3 protein.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 4058)
AUTHORS     Buchner,G., Bassi,M.T., Andolfi,G., Ballabio,A. and Franco,B.
TITLE       Identification of a novel homolog of the drosophila staufen protein
            in the chromosome 8q13-q21.1 region
JOURNAL     Genomics 62 (1), 113-118 (1999)
MEDLINE     20054364
REFERENCE   2 (bases 1 to 4058)
AUTHORS     Franco,B.
TITLE       Direct Submission
JOURNAL     Submitted (07-JUN-1999) B. Franco, TIGEM, Via Olgettina 58, 20132
            Milan, ITALY

FEATURES             Location/Qualifiers
     source           1..4058
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="teratocarcinoma-neuron"
                     /map="8q13-8q21.1"
                     202..1642
                     /codon_start=1
                     /product="39k3 protein"
                     /protein_id="CAB64341.1"
                     /db_xref="GI:6634457"
                     /translation="MLQINQMFSVQLSLGEQTWESESSIKKAQAQAVANKALTESTLP
KPVOKPSPKSNWNNGSITPTVELNGLAMKRGEPAIYRPLDPEPNRYRANYNFRGM
NQRYHCPVPKIFYVOLTVGNNEEFEGCKTRQARHNAAAKALQANEPIPERSPONG
ESGKDMDDDKDANKSEISLVEFIEALKRMNPVSFEVIKESGPBMKSFTVRVSGEFA
EEGENSKLSKKRAATTVLQELKLPLPVEKPKLFKKRKPTIVKAGPEYGQGMNP
ISRLAIOOAKEKEPDYVLISERGMPRRREFVMQVGNENAVGTGPNKIACKNNA
EAMLLOLGKASTNLQDLQLEKTENGWSPKPGFPETNTNTPKILHLSPDYQEME
ASRHKVISGTTLLGYLSPKDMNPSSSFSSISPTSNSATAIRELLMNGTSTAIAIGL
KGSSPTPCSPVPQPSKOLEYLARIQGFQV"

BASE COUNT        1280 a      731 c      737 g      1310 t
ORIGIN
alignment_scores:
Quality:          927.00              Length:          449
Ratio:            2.915                Gaps:            14
Percent Similarity: 70.824          Percent Identity: 48.775

alignment_block:
US-09-316-048-11 x HSA9062 ..

Align seg 1/1 to: HSA9062 from: 1 to: 4058

1 MetTyrlsProvalasProhisserArgmetGlnSerThrTyrserty 17
:::|||||:::|||||:::|||||:::|||:::|
439 ATCTACAGGCCATTAGATCCAAAGCCATTCCCAATATATAGAGCTAATTA 488
17 rglyMetarglglyAlatyrrproProarqtgtyrPhetyrrProphetrov 34
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
489 CAACTTTCGG..GGCATGTACAATCAGAGGTATCAT.....TGCCACAG 529
34 alProProleuleutyrgInValgluleuSerValglyglyGlnGlnphe 50
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
530 TCCTTAGATCTTTTAT...GTTACGCTCACCTGTAGGAATAATGAATTT 576
51 AsnglyLysglyLysmetargProProvalyShsaspalaProAlaar 67
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
577 TTGGGGAAGAAGACATCGACAAGCTGTAGACACAATGCCTGCAATGAA 626
67 galaleuArGthrleuGlInserrGluProleuProGluArgleuGluValA 84
:::|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```


PLPVEKPKLFFKKRPKTIWKAGPEYGGGMPISRLOIOQAKKEKEPDVLLSERGM
PARRREFVMQVKVGNENVAITGTPNKKIAKKNAEMLLQGYKASTNLQDLEKTEGEMK
GWSGPKGPEPTNPNPKILHLSPDVYOEMEARHKGISGTTIGYLSPKDMNOPSSS
FESISPTSNSATIARELIMGSTSTAEAIGLKGSSPTPCSPVQPSQOLEYLARIQG
FOALSAALKQFSEQGLPIDGAMNIEKGSLEKQAKHLREKADNNQAPPSIAQDCKKS
NSAV"

BASE COUNT 970 a 605 c 643 g 750 t
ORIGIN

alignment_scores:

Quality: 914.00 Length: 433
Ratio: 2.977 Gaps: 14
Percent Similarity: 70.901 Percent Identity: 49.423

alignment_block:

us-09-316-048-11 x AK002152 ..

Align seg 1/1 to: AK002152 from: 1 to: 2968

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTyr 17
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
588 ATCTACAGGCCATTTAGATCCAAAGCCATTCCTCAATATATAGAGCTAATTA 637
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPheProy 34
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
638 CAACCTTCGG...GGCATGTACAAATCAGAGCTATCAT...TGCCTCAG 678
34 aLProProLeuLeuTyrGlnValGlnLeuSerValGlyGlyGlnGlnPhe 50
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
679 TGCCTAAGATCTTTTAT...GTTCACTCACTGTAGAAATATGAAATTTT 725
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
726 TTTGGGAAGGAAGAACTCGACAAAGCTGTAGACACAAATGCTGCAATGAA 775
67 gAlaLeuArgThrLeuGlnSerGlnProLeuProGlnArgLeuGlnVala 84
:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
776 AGCCCTCCAAGCACCTGCAGAATGAACCTATTCCAGAAAGATCTCCTCAGA 825
84 snGlyArgGlnAla.....GlnGlnGlnAsnLeuAsnLysSer 96
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
826 ATGCTGAATCAGGAAGAGATATGATGATGACAAAGATGCAATATAGTCT 875
97 GlnLeuSerGlnValPheGlnLeuLeuAlaLeuLysArgAsnLeuProVala 113
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
876 GAGATCAGCTTAGTGTGAAATGCTCTGGAAGCGAAATATGCTGTGACG 925
113 nPheGlnValAlaArgGlnSerGlyProProHisMetLysAsnPheValT 130
:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
926 TTTTGAAGTTATTAAAGAAAGTGACACACATATGAAAGCTTTGTTTA 975
130 hrArgValSerValGlyGlnPheValGlyGlnGlyGlyLysSerLys 146
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
976 CTCGAGTGTACAGTAGAGAGTTCTCTGACAGAGGAAGAAATAGCAAA 1025
147 LysIleSerLysLysAsnAlaAlaArgAlaValLeuGlnLeuArgAr 163
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1026 AAACCTCTCCAAGAACGCGCTCGACACACCTCTTACAGAGCTTAAAAA 1075
163 gLeuProProLeuProAlaValGlnArgValLysProArgIleLysLysL 180
:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1076 ACTTCACACCTCTTCTGTGTGGAAAAAGCCAAACTATTTTAAAAAAC 1125
180 yssSerGlnProThrCysLysThrAlaProAspTyrGlyGlnGlyMetasn 196
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1126 GCCCTAAACAATAGTAAAGCGGACGACCAAGATATGCGCAAGGATGAAC 1175
197 proIleSerArgLeuAlaGlnIleGlnAlaLysLysGlnLysGlnPr 213
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1176 CCTATTAGCCGCTGGCGCAAAATTCAACAGAGCCCAAAAGGAAAGAGACC 1225
213 oGlnTyrMetLeuLeuThrGlnArgGlyLeuProArgArgArgGlnPhev 230

1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1226 GCATTATGTTTGTCTTTCAGAAAGAGGAATGCCCTCGACGTCGAGAAATTYG 1275
230 aImetGlnValLysValGlyHisIsthrAlaGlnGlyValGlyThrAsn 246
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1276 TGATGCAGGTGAAGGTAGGCAATGAAGTTGCTTACAGGAACAGGACCTTAAT 1325
247 LysLysValAlaLysArgAsnAlaAlaGlnAsnMetLeuGlnIleLeuGln 263
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1326 AAAAAGATAGCCAAAAAAATGCTGCAGAGCAATGCTGTACAACTTGG 1375
263 yPheLysValProGlnAlaGlnProAlaLysProAlaLeuLysSerGln 279
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1376 TTATAAA.....GCATCCACTAATCTTCAGAGATCAAC 1407
280 ..GlnLysThrProValLysLysProGlyAspGlyArgLysValThrPhe 295
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1408 TTGAGAAGACAGGGGAAACAAGAGATGAGTGGT..... 1442
296 PheGlnProSerProGlyAspGlnAsnGlyThrSerAsnLysAspGln 312
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1443CCAAAGCCTGGGTTTCTCTGAACCAACAATAAT..... 1475
312 uPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuP 329
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1476ACTCCAAAAGGAATCTTTC 1494
329 roMetValProGlnValAlaGlnAlaVal..... 338
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1495 ATTGTCTCTGATGTTTATCAAGAGATGGAAGCCAGCGCCACAAAGTA 1544
339GlyValSerGlnGlyHis..HisThrLysAspPheThrArgAl 352
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1545 ATCTCTGGCACTACTAGTACGCTATTGTGCACCCAAAGATATGAACCAACC 1594
352 aAlaProAsn.....ProAlaLysAlaThrValThrAlaMetI 365
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1595 TTCAAGCTCTTCTTCTCAGTATATCTCCACATCGAATAGTTCAGCTACAA 1644
365 leAlaArgGlnLeuLeuTyrGlyGlyThrSerProThrAlaGlnIle 381
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1645 TTGCCAGGGAACCTCTTATGAATGAACATCTTCTACAGCTGAAGCCATA 1694
382 ..LeuLysSerAsnIleSerSerGlyHisValProHisGlyProArgTh 397
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1695 GGTTTAAAGA.....AGTTCTCTACTCCCTGTTCTCCA...GT 1735
397 rArgProSerGlnLeuLeuTyrTyrLeuSerArgAlaGlnGlyPheGln 413
:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1736 ACAACCTTCAAAACAACCTGAATATATTAGCAAGGATTCAAGGCTTTCAG 1784

seq_name: gb_pr6:HSBB140E4

seq_documentation_block:

LOCUS HSB140E4 163253 bp DNA PRI 07-FEB-2000
DEFINITION Human DNA sequence from clone RP13-140E4 on chromosome X, complete
sequence.

ACCESSION AL121877
VERSION AL121877.13 GI:6933862

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 163253)
AUTHORS Heath,P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk
On Feb 7, 2000 this sequence version replaced gi:6911935.
During sequence assembly data is found these are annotated as variations
where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/ChrX>
RP13-140E4 is from the library RPC1-13.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pBACE3.6

This sequence is the entire insert of clone RP13-140E4.

FEATURES

source

```
1.163253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone_lib="RPC1-13.1"
25.188
repeat_region /note="L1M4 repeat: matches 5114. .5296 of consensus"
repeat_region /note="L1P3 repeat: matches 2841. .3072 of consensus"
repeat_region /note="L1P3 repeat: matches 1. .287 of consensus"
repeat_region /note="L1P3 repeat: matches 232. .2841 of consensus"
repeat_region /note="L1PA2 repeat: matches 3148. .6144 of consensus"
repeat_region /note="L1MD1 repeat: matches 5283. .6016 of consensus"
repeat_region /note="L1M1 repeat: matches 137. .290 of consensus"
repeat_region /note="L1MD2 repeat: matches 6016. .6330 of consensus"
repeat_region /note="L1M5 repeat: matches 7608. .7714 of consensus"
repeat_region /note="L1M5 repeat: matches 1. .305 of consensus"
repeat_region /note="L1M5 repeat: matches 7478. .7572 of consensus"
repeat_region /note="L1M5 repeat: matches 84. .305 of consensus"
repeat_region /note="L1M5 repeat: matches 7305. .7446 of consensus"
repeat_region /note="84 copies 2 mer tt 72% conserved"
repeat_region /note="L1M5 repeat: matches 8. .300 of consensus"
misc_feature /note="match: STS: Em:L41960; match: STS: Em:L41960"
repeat_region /note="L1M1 repeat: matches 2083. .2505 of consensus"
repeat_region /note="L1M1 repeat: matches 6004. .6303 of consensus"
repeat_region /note="L1M1 repeat: matches 1197. .2083 of consensus"
misc_feature /note="match: GSS: Em:AQ558019"
```

```
misc_feature 16840. .17277
/note="match: GSS: Em:B65742"
misc_feature complement(16845. .17241)
/note="match: GSS: Em:AQ149727"
misc_feature complement(16848. .17220)
/note="match: GSS: Em:AQ240745 Em:AQ474471"
repeat_region /note="MER4D repeat: matches 234. .973 of consensus"
repeat_region /note="L1P3 repeat: matches 5928. .6148 of consensus"
repeat_region /note="L1P3 repeat: matches 1. .554 of consensus"
repeat_region /note="MER41A repeat: matches 15. .198 of consensus"
misc_feature /note="match: GSS: Em:AQ017861"
repeat_region /note="MER57B repeat: matches 1. .403 of consensus"
repeat_region /note="MER57B internal repeat: matches 1. .806 of consensus"
repeat_region /note="MER57 internal repeat: matches 440. .2799 of consensus"
repeat_region /note="MER57 internal repeat: matches 3892. .4230 of consensus"
repeat_region /note="L1M5 repeat: matches 8. .286 of consensus"
repeat_region /note="MER57 internal repeat: matches 4230. .6114 of consensus"
repeat_region /note="L1P1 repeat: matches 5980. .6153 of consensus"
repeat_region /note="L1P1 repeat: matches 1. .304 of consensus"
repeat_region /note="L1P1 repeat: matches 5777. .5980 of consensus"
repeat_region /note="L1P1 repeat: matches 25672. .25672 of consensus"
repeat_region /note="L1P1 repeat: matches 26070. .26070 of consensus"
repeat_region /note="L1P1 repeat: matches 5397. .5790 of consensus"
repeat_region /note="MER57 internal repeat: matches 6489. .7537 of consensus"
repeat_region /note="MER57B repeat: matches 1. .114 of consensus"
repeat_region /note="MER57B repeat: matches 6763. .7631 of consensus"
repeat_region /note="L1M3 repeat: matches 28. .313 of consensus"
repeat_region /note="L1M3 repeat: matches 5376. .5993 of consensus"
repeat_region /note="L1M3 repeat: matches 4880. .5566 of consensus"
repeat_region /note="L1M4 repeat: matches 1. .301 of consensus"
repeat_region /note="L1M4 repeat: matches 3983. .6149 of consensus"
repeat_region /note="L1M4 repeat: matches 1. .440 of consensus"
repeat_region /note="L1M4 repeat: matches 2718. .3988 of consensus"
repeat_region /note="L1M4 repeat: matches 1. .369 of consensus"
repeat_region /note="L1M4 repeat: matches 1264. .2718 of consensus"
repeat_region /note="L1M4 repeat: matches 288. .1240 of consensus"
repeat_region /note="L1M4 repeat: matches 2349. .2404 of consensus"
repeat_region /note="L1M4 repeat: matches 59. .146 of consensus"
repeat_region /note="L1M4 repeat: matches 59. .146 of consensus"
```

```
/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region 41650. .41743
/note="47 copies 2 mer aa 62% conserved"
repeat_region 41744. .42318
/note="L1 repeat: matches 4553. .5128 of consensus"
repeat_region 42337. .42391
/note="Alu repeat: matches 251. .302 of consensus"
repeat_region 42360. .42395
/note="Alu repeat: matches 79. .114 of consensus"
repeat_region 42421. .43734
/note="L1 repeat: matches 4019. .5334 of consensus"
repeat_region 43735. .43834
/note="FLAM_C repeat: matches 32. .131 of consensus"
repeat_region 43835. .44621
/note="L1 repeat: matches 3232. .4019 of consensus"
repeat_region 44622. .44802
/note="L1M1 repeat: matches 6121. .6304 of consensus"
repeat_region 44803. .45815
/note="L1 repeat: matches 2126. .3232 of consensus"
repeat_region 45514. .46762
/note="L1M1 repeat: matches 1155. .2794 of consensus"
repeat_region 46760. .47327
/note="L1M1 repeat: matches 59. .642 of consensus"
repeat_region 47325. .48451
/note="L1M1 repeat: matches -1334. -.154 of consensus"
repeat_region 48533. .48727
/note="L1 repeat: matches 4030. .4226 of consensus"
repeat_region 48717. .49410
/note="L1M1 repeat: matches 5092. .5805 of consensus"
repeat_region 49411. .49506
/note="48 copies 2 mer at 69% conserved"
repeat_region 49528. .50023
/note="L1M1 repeat: matches 5787. .6308 of consensus"
repeat_region 50050. .50333
/note="Alu repeat: matches 1. .300 of consensus"
```

```
alignment_scores:
  Quality: 705.50      Length: 446
  Ratio: 2.383        Gaps: 15
Percent Similarity: 66.368      Percent Identity: 43.722
```

alignment_block:

US-09-316-048-11 x HSB140E4 ..

Align seg 1/1 to: HSB140E4 from: 1 to: 163253

```
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
   ::::::::::::::::::::
159120 ATCCACAGGCCATTAGATCCAAAGCCATTCCCAATATATAGACCTAATTA 159169
   ::::::::::::::::::::
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
   ::::::::::::::::::::
159170 CAACCTTTCAA...GTCATGTATATATCAGAGGTATCATTTGCCCAATA... 159212
   ::::::::::::::::::::
34 alProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50
   ::::::::::::::::::::
159213 ....CCTAAGATCTTTTATGTTCAGCTCAT.GTAGGAATATATGAATTT 159256
   ::::::::::::::::::::
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
   ::::::::::::::::::::
159257 TTGGGGAGGAAAGCC.....AGACACAATGCTGCAATGAA 159293
   ::::::::::::::::::::
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
   ::::::::::::::::::::
159294 AGACATCCAAAGCACTGCAGATGTAACCTATTCCAGAAAGATCGCCTTAGA 159343
   ::::::::::::::::::::
84 sngLysArgGluAla.....GluGluGluAsnLeuAsnLysSer 96
   ::::::::::::::::::::
159344 ATGGTGAATTCAGAAAGGATATATGATGACAAAGATGCCAATAATAGCCT 159393
   ::::::::::::::::::::
97 GluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAs 113
   ::::::::::::::::::::
159394 GAGATCAGCTTATGTGTTGAATATGCTCTGAACTGAATATATACCTGTAG 159443
   ::::::::::::::::::::
```

```
113 nPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheValT 130
   ::::::::::::::::::::
159444 TTCTGAGGTATTTAAAGAAAGTGACCAACCATATATGAAAGCTTTGTTA 159493
   ::::::::::::::::::::
130 hArgValSerValGlyGluPheValGlyGluGlyGlyLysSerLys 146
   ::::::::::::::::::::
159494 CTCGAGTGTAGTAGGAAGTTCTCTACAGAAAGAGGAAGAAATATAGCAAA 159543
   ::::::::::::::::::::
147 LysIleSerLysLysAsnAlaAlaArgAlaValLeuGlnLeuArgAr 163
   ::::::::::::::::::::
159544 AAACCTCCAGAAAGCGCTGTCACACCAACCGCTTACAGAGCTTAAAAA 159593
   ::::::::::::::::::::
163 gLeuProProLeuProAlaValGluArgValLysProArgIleLysLysL 180
   ::::::::::::::::::::
159594 ACTTCCACCCTTCCTGCTGTGGAAGCAAAATATATTTTAAAAAAC 159643
   ::::::::::::::::::::
180 ySSerGlnProThrCysLysThrAlaProAspTyrGlyGlnLysMetAsn 196
   ::::::::::::::::::::
159644 GCCCTAAACAATAGTCAAGGCCGACCAAGATA.GGCCAAGGATGAAC 159692
   ::::::::::::::::::::
197 ProIleSerArgLeuAlaGlnIleGlnAlaLysLysGluLysGluPr 213
   ::::::::::::::::::::
159693 CCTATTAGCTGCTGCGCCAAATTCACAGGCCCAAAAGGAAGAGAGCC 159742
   ::::::::::::::::::::
213 oGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheV 230
   ::::::::::::::::::::
159743 AGATTATGTTTGTCTTCAGAAAGAGGAAATGCCGTGATGTGAGAATTTG 159792
   ::::::::::::::::::::
230 alMetGlnValLysValGlyHisThrAlaGluGlyValGlyThrAsn 246
   ::::::::::::::::::::
159793 TGATGTATATGAGGTAGGCAATGAGCTGCTACAGAACCGGACCTTAAT 159842
   ::::::::::::::::::::
247 LysLysVal...AlaLysArgAsnAlaAlaGluAsnMetLeuGluIleL 262
   ::::::::::::::::::::
159843 AAAAAGGCCAAAAAAGAGTTGCAAGTAGCAATGCTATTACAAAC 159892
   ::::::::::::::::::::
262 euGlyPheLysValProGlnAlaGlnProAlaLysProAlaLeuLysSer 278
   ::::::::::::::::::::
159893 TTGGTTATATAA.....GCATCCATTATCTTCAGAT 159924
   ::::::::::::::::::::
279 Glu...GluLysThrProValLysLysProGlyAspGlyArgLysValTh 294
   ::::::::::::::::::::
159925 CAACCTGAGAGACA.....GGGGAACAAGAGG 159953
   ::::::::::::::::::::
294 rPhePheGluProSerProGlyAspGluAsnGlyThrSerAsnLysAspG 311
   ::::::::::::::::::::
159954 ATGAGTGAATCCAAAGCGCTGGTTCTCTGAACCAACAATTAAT..... 159996
   ::::::::::::::::::::
311 LuGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIle 327
   ::::::::::::::::::::
159997 .....ACTCAAAAGGAATT 160011
   ::::::::::::::::::::
328 LeuProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyH 344
   ::::::::::::::::::::
160012 CTTCATTTGTCCTCGATGTTATCAAGAGATGGAAGCCAGACCAACAA 160061
   ::::::::::::::::::::
344 shiThrLysAspPheThrArgAlaAlaProAsnPro..... 356
   ::::::::::::::::::::
160062 AGTAACCTCTGGCACCTACTCTAGGCTATTGTGCACCCAGAGATATGAAC 160111
   ::::::::::::::::::::
357 .....AlaLysAlaThrValThraLa 363
   ::::::::::::::::::::
160112 AACCTTCAAGCTCTTGTTCAGTATGATATCCACATCAAAATAGTTACAG 160161
   ::::::::::::::::::::
364 MetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGlu.. 379
   ::::::::::::::::::::
160162 ACAATGGCCAGGAACCTCTTATGAATGGAACATCTTCTACAGCTGAAGC 160211
   ::::::::::::::::::::
380 .ThrIleLeuLysSerAsnIleSerSerGlyHisValProHisGlyProA 396
   ::::::::::::::::::::
160212 CACAGGTTTAAAGA.....AGTCTCTCTACTCCCCCTGTCTCTCCA. 160254
   ::::::::::::::::::::
```

396 rgThrArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPhe 412
 :::::::::::::::
 160255 ..GTACACCTTCACAAACAACACTGGAATATTTAGCAAGGATTCACAGGCTTT 160302
 413 GlnValGluTyrLysAspPheProLysAsnAsnLys 424
 :::::::::::::::
 160303 CAGGTATGAATTAAAGC.....AAAACAAAAA 160332

seq_name: gb_pr4:AL133174

seq_documentation_block:

LOCUS AL133174 113836 bp DNA PRI 14-JUN-2000
 DEFINITION Human DNA sequence from clone RP3-470L14 on chromosome 20, complete
 sequence.
 ACCESSION AL133174
 VERSION AL133174.15 GI:8573761
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 113836)
 AUTHORS Ramsay,H.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jun 20, 2000 this sequence version replaced gi:8388434.

COMMENT

During sequence assembly data is compared from overlapping clones.
 where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone configs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 RP3-470L14 is from the library RPCI-3 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: PCYPAC2
 IMPORTANT: This sequence is not the entire insert of clone
 RP3-470L14 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true right end of clone RP3-470L14 is at 113836 in this
 sequence. The true right end of clone RP1-15566 is at 100 in this
 sequence.

FEATURES

Source 1. 113836
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP3-470L14"
 /clone_lib="RPCI-3"
 3061. 3077
 /note="Weak data"
 misc_feature 95583. 95627
 /note="Single clone region. pUC subcloned to give 12x
 coverage.. Assembly confirmed by restriction digest data."
 BASE COUNT 33974 a 25115 c 24233 g 30514 t
 ORIGIN

alignment_scores:
 Quality: 701.00 Length: 429
 Ratio: 4.124 Gaps: 3
 Percent Similarity: 39.627 Percent Identity: 37.762

alignment_block:

US-09-316-048-11 x AL133174/rev ..

Align seg 1/1 to reverse of: AL133174 from: 1 to: 113836

281 LysThrProValLysLysProGlyAspGlyArgGlyValThrPhePheG1 297
 :::::::::::::::
 38483 AAGACACCCATTAAGAAACCAAGGGGATGGAAGAAAGTAACCTTTTGA 38434
 297 uProSerProGlyAspGluAsnGlyThr..... 306
 :::::::::::::::
 38433 ACCTGGCTCTGGGATGAATAATGGGACTAGTGTGATCTTAATATCG 38384
 306 306
 38383 TCCTCCTCTGAACGTGATTCTGGGATTTCAGAGGCTCTGAATAACAGCA 38334
 306 306
 38333 GGAGCCACAGCCAGTGAGGTCCCGCATGCAAAAGTCTTATCATTCCTG 38284
 306 306
 38283 ATTCCCTGCGCTGCCATGGCTGCCAAGGGAGGGCCAGCGTGGGTTGA 38234
 306 306
 38233 TGGTGTCATTGCGTTTTCACAGGCTGTGATGTGATTCGTGTTACCGT 38184
 307Ser..AsnLysAspGluGlnPheArgMetProTyrLeu 318
 :::::::::::::::
 38183 GTATGTCCCTCCACAGTAATAAGAGATGAGTTCAGGATGCTTATCTA 38134
 319 SerHisGlnGlnLeuProAlaGlyTleuProMetValProGluValAl 335
 :::::::::::::::
 38133 AGTCATCAGCAGAGCTGCGCTGCTGGAATCTTCCCATGGTGCCGAGTCCG 38084
 335 agLAlaValGlyValSerGlnGlyHisHisThrLysAspPheThArga 352
 :::::::::::::::
 38083 CCAGGCTGTAGGAGTTAGTCAAGGACATCACACCAAGATTTTACCAGGG 38034
 352 laAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 368
 :::::::::::::::
 38033 CAGCTCCGAATCCTGCCAAGGCCACGGTAACCTGCCATGATAGCCCGAGAG 37984
 369 LeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSerAs 385
 :::::::::::::::
 37983 TTGTTGATGGGGGCACCTGCGCCACAGCCGAGACCATTTTAAAGAATAA 37934
 385 nIleSerSerGlyHisValProHisGlyProArgThrArgProSerGluG 402
 :::::::::::::::
 37933 CATCTCTCAGGCCACGATACCCCATGGAACCTCTCAGCAGACCCCTTGAGC 37884
 402 InLeuTyrTyrLeuSerArgAlaGlnGlyPheGln..... 413
 :::::::::::::::
 37883 AACTGGACTATCTTTCAGAGTCCAGGGATTCAGAGTAACGTCTGGCCT 37834
 413 413
 37833 GAGCTGTGATGGCTGTGCTCAGGACAGTCATCTTTGGCTTCTCTCAGA 37784
 413 413
 37783 AAATGAACCTTGAACTTGTACTTTTCTTAATCAAGTCTGTCCAGG 37734
 413 413

37733	GATGGCTGGGAGCAGTGCATGCAAGCGGTAGGAGGTATAGAAGTATTGTA	37684
413	413
37683	AGCATATGTGGTAAGAGAAATATCTTTATCCTTTGGTAGTAAGATTGCA	37634
413	413
37633	CATGCAAGCCAAACCCTAAAGCTTCTAACCACCTAGCAGTGGTTGAATP	37584
413	413
37583	TGATTTCAAGAACAAGACTCATCGTGTGACACTTCTGTGTGGTAGGA	37534
413	413
37533	AAGGATATGGCTTCTCCAGTCTTGGGCTTCTGAGCAGACACCTAGGTC	37484
413	413
37483	CTGGAAGGTTCTGGGCTGACAGGGCAGGTGTTAGTATGGCCACAGTGGG	37434
413	413
37433	CTGAGTTGGAAAAAAGAGCTGTGATCTTGTGTGGGGCTGAGCTGC	37384
413	413
37383	AGCCCTCCAGCCCACTCACTACTGTGTTCCCTACCTAGCCTCAGCTCTC	37334
414ValGluTyrLysAspPheProLysAsnAsnLysAsnG	426
37333	TTTTCCCCCTTAGGTGAATACAAAGACTTCCCCCAAAAACACACAGAAGC	37284
426	LucValSerLeuIleAsnCyssSerSerGlnProProLeuValSerHis	442
37283	AATTGTATCTCTTATCAATTGCTCTCCTCAGCCACCTCTGTATCAGCCAT	37234
443	GLYIleGLYlysAspValGluSerCysHisAspMet	454
37233	GGTATCGCAAGGATGTGAGTCTCGCCATGATATG	37198

```

seq_name: gb_in3:DROSTAUFEIN
seq_documentation_block:
LOCUS      DROSTAUFEIN      5237 bp      mRNA      INV      26-APR-1993
DEFINITION D.melanogaster mRNA, complete cds.
ACCESSION  M69111
VERSION    M69111.1  GI:158505
KEYWORDS
SOURCE
ORGANISM
. Drosophila melanogaster 0-4 hour embryo cDNA to mRNA.
  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 5237)
Johnston,D.S., Beuchle,D. and Nusslein-Volhard,C.
Staufen, a gene required to localize maternal RNAs in the
drosophila egg
Cell 66, 51-53 (1991)
91300552

FEATURES
      source
          location/Qualifiers
              1..5237
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /dev_stage="0-4 hour embryo"
                <1..5237
                275..3355
                /codon_start=1
                /protein_id="AA073062.1"
                /db_xref="GI:158506"
                /translation="MQHNVEARPPAAPHIRAAHHHSHSHAHMHILHPGMEQHLG

```

QQPPPPQQPPHRDLHARLNHHHLHAQOQQQQQQQTSSNQAAVAAGAAYHHGNIINSNS
GSIINSNSNQOMOKIRQOHQLSSSLGNQPPGPPPOAFENPLAGNPALAYNOLPPH
PPHMAAHLGSIYAAPPPHYYSQAKPAKYNHYGSNANSNSNGSNNSNYAPKAILONT
YRNOKVVPVVOEVTVPPEPVPYTNATNTSTNSNSTNSIASEPYTOEDTSOKPETROE
PASADHDVSTGNIDATGALSNEDTSSSGRGKDKTPMCVLNELARYNKITHQYRLTEE
RGPACHCKFTVTMLGDEEYSADGEFKKAQHLAASKAIETEMKHPPKIRRSEEG
PMRTHITPTVELNALAMKLGRQFYLLDPTQIPPTDSIAPPEAGGHLPTLPGPMVO
PPPPPAYALRQRLNGFVPIPSQPMHPHFHGGQRFPFKFPRFALLPPLAGHVHH
GPNCPFPSPVPTPSPKITLFWGKQKQVIGRTLOQAKHDAARALOVLKTOAISASEEA
LEDSMDEGDKKSPIQSQVHEIGIKRNMVHFVLRREGPAHMKNFITACIVGSIVTEGE
GNCKRYSKKRAAEKMLVELQKLPPLTPKQIPLKRIKVTTPKSGAAREGSVSGTD
GTMQIGKPERKKRLNPPKDKLIDMDADANPITKLIQLOQTRKEKPIFELIAKNGNET
ARRREFVMEVSAGSTARGTGNSKLAKRNAAOALFELLEAVOYTPNETOSSSEECT
SATMSAVTAPAVEATAEGKVPMNAVTPVGPMPGILIRQNKPKPAKKRQOIVIVKSNEVS
KEEANKKEVAVAEENSNSNANSNGSSSSGDSQATEAASESALNTSTGSNTSGVSS
NSSNVGANTDGNNAHESKNNTESSNSTNTQSAGVHMKEQLLYLSKLLDFEVNFSDY
PKGNHNEFLTIVLTSTHPQICHGVKSSSEESQNDASNALKILSKLGLNNAMK"

```

alignment_scores:          Quality: 634.50      Length: 575
                          Ratio: 1.995       Gaps: 14
Percent Similarity: 55.304   Percent Identity: 30.435

Alignment_block:
US-09-316-048-11 x DROSTAUFEN ..

Align seg 1/1 to: DROSTAUFEN from: 1 to: 5237

3 LysProValAspProHisSerArgMetGlnSerThrTyrSerTyrglyme 19
:::||||: ||||| ::: |||
1691 CAGCCGATGCATCCGCAT.....TTCTTCATGGACC 1722

19 targlyglYAlaTyrProProAArgTyrPheTyrProPheProValProp 36
:::|||||::: ||| ::|||
1723 TGGTCAGCGACCGTTTCCTCCGAAGTTTCCATCAGCGCTTGCGCTGCCAC 1772

36 roLeuleu..... 38
|| |||
1773 CGCCTCTCGAGGCTCATGTACATCATGGACCGGAATGGCCCATTTCCATCC 1822

39 .....TyrGlnValGluleuSerValGlygInGlnPh 50
::::: || | |||| |:::|
1823 GTGCCTACGCCACCACCAAAATCACCCCTGTTGTGGGCAAGCAGAAGTT 1872

50 eaSnGLyLysGLyLysMeTarGProProValLysHisAsPaLaProAlaa 67
| ||| |||: :||| ||||| |||
1873 TGtGGGCATCGAGCGCACACTGCAGCAAGCGAAACATGACCGTGCGGCGA 1922

67 rGaLaLeuArGThrLeucInSerGluProleu.....ProGUArg 80
|||||:::|||||::: |||
1923 GAGCGCTGCAGGTTCTCAGACACAGGCCATATCTGCCTCTGAGGAGGCT 1972

81 LeuGLuValasngLyArgGluAlagLuGLuasnLeuasnLysSergl 97
||||| ::::: ::|||: |||||
1973 CTCGAG.....GATTGATGATGAGGCTGACAAGAAGTCGCC 2010

97 uIleSerGlnValPheGluIleAlaleuLysArgAsnLeuProValasp 114
||||| ||||| ::::: |||||: |||:|
2011 CATCTCGCAGGTTTCACGAGATCGGAATTAAAGCGCAACATGACCGTTCAT 2060

114 heGLuValAlaArgGluSerGlyProProHISmetLysasnPheValThr 130
|:::| |||::: ||||| | |||||:::|
2061 TTAAGGTGTTGCGCGAGAGGGGCGCGCACATGAAGACTTATAACA 2110

131 ArgValSerValGlyGluPheValGlyGluGlyLysSerLysly 147
||| |::: ||| ||||| |||||::: |||
2111 GCCTGcATtGTGGGATCTATTGTcACCGAAGAGAGAAATGcCAAAAA 2160

147 sILeserLysAsnAlaAlaArgAlaValLeuGluInLeuArgArgL 164
|:::| |||||::: |||||: ||| |||:::|

```



```

2161 AGTGTCCAAGAACGCTGCCGCGGAAAAAGATGCTCTGTCGAGTTGCCAAAAGC 2210
164 euProProLeuProAlaValGluArgVal..... 173
|||||
2211 TGCCTCCCTTACGCGCCACCAAGACAGCCCATTTGAAGCGATTAAAGTT 2260
173 ..... 173
2261 AAGACACCAGCAAAAGTGTGCAGCTGCCCGAGAAGTTCCGTTGTGTC 2310
174 .....LysProArgIleLysLysLys 181
|||||
2311 GGGCAGCTGATGGTACTATGCACAAACGGGAAAGCCCGGAGACGATAACGCC 2360
181 erGlnProThrCysLysThrAlaProAspTyrGlyGlnGlyMetAsnPro 197
|||||
2361 TCATTCGCCCGAAGACAACTGATAGACATGACGATGCCGATAATCCC 2410
198 lIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProG1 214
|||||
2411 ATTACAAAGCTAATTCAACTGCACAGACGCGTTAAGGAAAAAGAACCCAT 2460
214 uTyrMetLeuLeuThrGluArgGly.....LeuProArgArgArgGluP 229
|||||
2461 TTTCGAGCTGATAGCCAAAACGGCAATGAACCCGCTCGACAGCGGAAT 2510
229 heValMetGlnValLysValGlyHisThrAlaGluGlyValGlyThr 245
|||||
2511 TCGTTATGGAGGTCTCCGCCAGCGGAGACACAGCCCGTGGAACAGGCAAC 2560
246 AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLe 262
|||||
2561 AGCAAGAAATTGGCAAAAGCAAAATGCTGCACAGGCTCTATTGAACTGCT 2610
262 uGlyPheLysValProGlnAlaGlnProAlaLysProAlaLeuLysSerG 279
|||||
2611 G.....GAAGCTGTTCAAGTCAACCAACCAACCAACACAGTCATCGG 2654
279 lUGlu..... 280
|||||
2655 AGGAATGCTGCACTTCGGCAACAATGTCCGCGTTACAGCGCCCGCAGTG 2704
281 .....LysThrProVal..... 284
|||||
2705 GAGGCAACTGCGGAAGCAAGTGCCCATGTGTAACAAGCCAGTGGTCC 2754
285 .....LysLysProGlyAspGlyA 291
|||||
2755 CATGCCAGGCATTTTAATACTGCGCCAGAACAAAGAAACCAAGAAAGA 2804
291 rg..LysValThrPhePheGluProSerProGlyAspGluAsnGlyThr 306
|||||
2805 GGGATCAAAATTGTAAATTGTCAAGTCTAACGTGGAATCCAAGAGAGAGAG 2854
307 SerAsnLysAspGluGluPheArgMetProTyrLeuSerHisGlnGlnLe 323
|||||
2855 GCCAAATAAGAGGTGGCAGTTCAGCTGAGAGAAATAGCAACAACAGTGC 2904
323 uProAlaGlyIleLeuProMetValProGluValAlaGlnAlaValGlyV 340
|||||
2905 AAACCTCTGGC.....GACAGCAGCAACAGCAGCAGCGCGG 2939
340 aISerGlnGlyHisHisThrLysAspPheThrArgAlaAlaProAsnPro 356
|||||
2940 ATTCGCAAGCC.....ACAGAGCGCGCTAGTGAGAGT 2971
357 AlAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyG1 373
|||||
2972 GCACTGAATACCTCAACCGGCGACAATACG.....AGTGG 3006
373 yThrSerProThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyH 390
|||||
3007 TGTGAGTAGCAACAGCAGCAATGTTGGAGCCAAACACAGCATGTGTAACAACC 3056

```

```

390 iSValProHisGlyProArgThrArgProSer..... 400
|||||
3057 ATGCCGAGAGTAAACAAACACGAGAGATAGCAGCAACAGTACGAGTAAC 3106
401 .....GluGlnLeuTyrTyrLeuSerAr 408
|||||
3107 ACACAAAGCGCTGAGTGCACATGAAGAGAGCAGCTCTTGTACCTTAGTAA 3156
408 galAGlnGlyPheGlnValGluTyrLysAspPheProLysAsnLysA 425
|||||
3157 ACTGCTGACTTTGAGGTCAACTTCTCGACTATCCGAAAGCAATCACA 3206
425 snGluCysValSerLeuIleAsnCysSerSerGlnProProLeuValSer 441
|||||
3207 ACGAGTTCCTGACCATCGTGACATTGTCCACACATCCGCCGCAAGATTGCG 3256
442 HisGlyIleGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAs 458
|||||
3257 CATGCCGTTGGCAAGAGCTCCGAGAGATCGCAGAATGATGCTGCAAGCAA 3306
458 nIleLeuLysLeuLeuSerGluLeu 466
|||||
3307 TGCCTTGAAATCTCTCAGCAAGCTG 3331

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:28:38 ; Search time 701.87 Seconds
(without alignments)
1721.839 Million cell updates/sec

Title: US-09-316-048-3
Perfect score: 3217
Sequence: 1 acttcctgcgcggcgtcgcg.....taatactaaaaaaaaa 3217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*

1: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3213.8	99.9	3217	21	Z93323	Human stauflen CDNA
2	3131.4	97.3	3190	20	X90786	DNA encoding human
3	3053.8	94.9	3142	21	Z93324	Human stauflen CDNA
4	3037.4	94.4	3260	21	Z93327	Human stauflen CDNA
5	2914.8	90.6	3506	21	Z93322	Human stauflen CDNA
6	1273	39.6	2859	21	Z93328	Mouse stauflen CDNA
7	1107.2	34.4	1127	20	X39991	Prostate cancer as
8	798.8	24.8	946	20	X40002	Prostate cancer as
9	755.6	23.5	844	20	X40000	Prostate cancer as
10	753.2	23.4	773	20	X39992	Prostate cancer as
11	551	17.1	825	20	X40001	Prostate cancer as
12	501.4	15.6	794	20	X39993	Prostate cancer as

C	13	433.8	13.5	660	21	Z79981	Human colon cancer
	14	412.2	12.8	727	21	Z80269	Human colon cancer
C	15	327.8	10.2	708	21	Z80703	Human colon cancer
	16	286.6	8.9	305	16	T20075	Human gene express
	17	126.6	3.9	300	20	Z17757	Human gene express
	18	126.6	3.9	300	20	Z14226	Human gene express
	19	122	3.8	769	20	Z17411	Human gene express
	20	83	2.6	767	20	Z16120	Human gene express
C	21	55	1.7	1000	21	A02484	Human colon cancer
	22	52	1.6	12001	16	Q76213	HSV L/ST region.
	23	51.8	1.6	2188	20	Z77506	Human ovarian tumo
	24	51.6	1.6	936	20	X56375	Human DNA-dependen
	25	51.2	1.6	53526	19	T94101	Human PKD1 gene.
	26	51.2	1.6	53577	17	T18551	Human polycystic k
	27	51.2	1.6	53577	19	T94108	Human PKD1 locus b
	28	50.4	1.6	10732	21	A10594	Gene encoding a su
	29	49	1.5	1337	20	Z17263	Human gene express
	30	48.8	1.5	867	21	Z98056	Human secreted pro
C	31	48.6	1.5	15899	19	V62164	HSV-2 strain SB5 C
	32	48.6	1.5	16812	19	V62175	HSV-2 strain SB5 C
C	33	48.4	1.5	38186	20	Z32028	Human METH1 relate
	34	48	1.5	8670	20	Z32059	Human METH2 relate
C	35	47.4	1.5	3198	20	X02974	Human IL-1ra BAC c
	36	47	1.5	1235	20	Z16147	Human gene express
	37	47	1.5	23142	21	A35158	Human adenosine re
	38	47	1.5	30417	21	A35160	Human adenosine re
C	39	45.8	1.4	1446	20	Z16805	Human gene express
	40	45.4	1.4	80240	20	V83940	NC-contig derived
	41	45.4	1.4	80595	20	V83939	HC-contig derived
	42	45	1.4	4897	11	Q03259	Pseudorabies virus
C	43	44.8	1.4	1102	20	Z16923	Human gene express
	44	44.8	1.4	1725	12	Q11243	PRV 28k gene. pse
C	45	44.6	1.4	1558	12	Q10896	Encodes Xenopus Bo

ALIGNMENTS

RESULT 1	
ID Z93323	standard; cDNA; 3217 BP.
AC Z93323;	
DT 04-JUL-2000	(first entry)
XX	
DE	Human stauflen CDNA.
XX	
KW	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW	acquired immune deficiency syndrome; protease; human; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	363..1853
FT	/*tag= a
XX	/product= Staufen
PN	CA2238656-A1.
XX	
PD	22-NOV-1999.
XX	
PF	22-MAY-1998; 98CA-2238656.
XX	
PR	22-MAY-1998; 98CA-2238656.
XX	
PA	(UYMO-) UNIV MONTREAL.
XX	
XX	Duchaine T, Wickham L, Cohen EA, Luo M, Desgrosellers L;
PI	Mouland AJ;
XX	
DR	WPI; 2000-246924/22.

DR P-PSDB; Y83108.

PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful for treating retroviral infections especially human immunodeficiency virus infections (HIV)

PS Claim 4; Figure 1a; 96pp; English.

Staufen is a RNA binding protein which interacts with double stranded RNA and/or rough endoplasmic reticulum. It is involved in mRNA transport and localization. Mammalian SP contain multiple ds RNA-binding domains. Chimeric proteins comprising portions of staufen with proteins with RNase or protease activity can be used to treat virus infections. The RNase or protease activity of the fusion protein prevents proper maturation of the virus. Mammalian staufen proteins seem to recognise double stranded RNA structure rather than any sequence specific position. 2 bacterially-expressed fusion proteins used in an RNA-binding assay (his/hsp (human SP) and MBP/MSF (murine SP)) both proteins strongly bound double stranded RNA. Both fusion proteins also directly bound labeled ds RNAs and RNA/DNA hybrids. Staufen proteins may therefore be useful in the treatment of RNA virus infections, especially those caused by retroviruses, in particular human immunodeficiency virus. This sequence is an alternatively spliced sequence to the one given in 2933322.

5Q Sequence 3217 BP; 889 A; 796 C; 742 G; 790 T; 0 other;

Query Match	99.98%	Score 3213.8;	DB 21;	Length 3217;
Best Local Similarity	99.98%	Pred. No. 0;		
Matches 3215; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

[illegible]

QY	601	agaaatgagcccccctgcagagaagcctggaagtgtaatggaagaaatccgaagaagaaaaatc	660
Db	601	agaatgagccccctgcagagaagcctggaagtgtaatggaagaaatccgaagaagaaaaatc	660
QY	661	tcaataaattcgaataaagtcaagtgtttgagattgcaattaaacggaacttgcctgtga	720
Db	661	tcaataaattcgaataaagtcaagtgtttgagattgcaattaaacggaacttgcctgtga	720
QY	721	atttcaggttgcgccggagagatgccccaccacatgaagaacttgttgaccaagttt	780
Db	721	atttcaggttgcgccggagagatgccccaccacatgaagaacttgttgaccaagttt	780
QY	781	cgtgtgagggaagtttgttgggggaaagtgaaagggaaaaagcaagaagatttcaaaagaaaaatg	840
Db	781	cgtgtgagggaagtttgttgggggaaagtgaaagggaaaaagcaagaagatttcaaaagaaaaatg	840
QY	841	ccgcataagctgttctcttgagggagctgaaagaattaccgccccctgcctgcagttgaaacgag	900
Db	841	ccgcataagctgttctcttgagggagctgaaagaattaccgccccctgcctgcagttgaaacgag	900
QY	901	taaagcctagaatcacaagaaagaaacaaaccatagttcaagccacagacaagccccagaat	960
Db	901	taaagcctagaatcacaagaaagaaacaaaccatagttcaagccacagacaagccccagaat	960
QY	961	atggcaggggatacaatccgattagccgactgcccagatccagcagcagcaaaaaagaga	1020
Db	961	atggcaggggatacaatccgattagccgactgcccagatccagcagcagcaaaaaagaga	1020
QY	1021	aggagccagagtacacgctctctcacagagcgagcgctcccgcgccgagggagttgtga	1080
Db	1021	aggagccagagtacacgctctctcacagagcgagcgctcccgcgccgagggagttgtga	1080
QY	1081	tgcaggtgaaagtttgnaaaccaactgcagaaggaacgggcaaccaagaaggtggcca	1140
Db	1081	tgcaggtgaaagtttgnaaaccaactgcagaaggaacgggcaaccaagaaggtggcca	1140
QY	1141	agcgcaatgcagccgagaacatgctgagatccttggttcaaaagtcccgcagcggcagc	1200
Db	1141	agcgcaatgcagccgagaacatgctgagatccttggttcaaaagtcccgcagcggcagc	1200
QY	1201	ccaaccaaacccgcactcaagtcagaggagaagacacccaataaagaaacaggggagtga	1260
Db	1201	ccaaccaaacccgcactcaagtcagaggagaagacacccaataaagaaacaggggagtga	1260
QY	1261	gaaaagtaaccttttttgaaacctgcctgtgggatgaaatgggactagtaataaagag	1320
Db	1261	gaaaagtaaccttttttgaaacctgcctgtgggatgaaatgggactagtaataaagag	1320
QY	1321	atgagttcaggatgccttatctaagtcatacagcagctgcctgtgaaattcttcccatgg	1380
Db	1321	atgagttcaggatgccttatctaagtcatacagcagctgcctgtgaaattcttcccatgg	1380
QY	1381	tgcgccaggtcgcccaagcgtgtagagttagttcaagggatcacaccaagaatttaacca	1440
Db	1381	tgcgccaggtcgcccaagcgtgtagagttagttcaagggatcacaccaagaatttaacca	1440
QY	1441	gggcagctccgaatcctctgcacaaggccaaggtaaactgcatagtagcccgagagttgtgt	1500
Db	1441	gggcagctccgaatcctctgcacaaggccaaggtaaactgcatagtagcccgagagttgtgt	1500
QY	1501	atgggggacacctgcgccacagccgagaccattttaagaataacatctctcagggcacg	1560
Db	1501	atgggggacacctgcgccacagccgagaccattttaagaataacatctctcagggcacg	1560
QY	1561	taccocatgacctctcacagagaccctctgagcaactgactatcttccagagttccagg	1620
Db	1561	taccocatgacctctcacagagaccctctgagcaactgactatcttccagagttccagg	1620
QY	1621	gattccaggttgataatacaagaacttccccaaaaacaacaagaagaatttgtatctctta	1680
Db	1621	gattccaggttgataatacaagaacttccccaaaaacaacaagaagaatttgtatctctta	1680
QY	1681	tcaattgtctctctcagccacctctgatcagcccatggtatcggcaagagatgtggaatcct	1740

Db 1681 tcaattgctcctctcagccacctctgatacagccagtgatcgcaaggaatgtgagtcct 1740
 QY 1741 gccatgatatgtgctgcgcgaacatcttaagtgtctgtgagtttgaccaacaagta 1800
 Db 1741 gccatgatatgtgctgcgcgaacatcttaagtgtctgtgagtttgaccaacaagta 1800
 QY 1801 cagaagatgccaaagaaacaggaacacgaacatgtctgtgtggaagtgctgaaccttt 1860
 Db 1801 cagaagatgccaaagaaacaggaacacgaacatgtctgtgtggaagtgctgaaccttt 1860
 QY 1861 ctggccatgaaccattataaaaatcccaacatataactgaaataactgaaactgttga 1920
 Db 1861 ctggccatgaaccattataaaaatcccaacatataactgaaataactgaaactgttga 1920
 QY 1921 aaatttgaatttctgatacctccagtggtggtgagacacagtggtgaaagatgtgg 1980
 Db 1921 aaatttgaatttctgatacctccagtggtggtgagacacagtggtgaaagatgtgg 1980
 QY 1981 cagcagcaggggaagacacacagaaacacacaggaagcggtgtggtggtgactgtgtg 2040
 Db 1981 cagcagcaggggaagacacacagaaacacacaggaagcggtgtggtggtgactgtgtg 2040
 QY 2041 ggggttgtgtgtgatagcacctcggtgacactgtggtggtggtggtggtggtggtg 2100
 Db 2041 ggggttgtgtgtgatagcacctcggtgacactgtggtggtggtggtggtggtggtg 2100
 QY 2101 gggggaagaagggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 2160
 Db 2101 gggggaagaagggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 2160
 QY 2161 ggcacctcgt 2220
 Db 2161 ggcacctcgt 2220
 QY 2221 tctcaagctgtgctcactcagacacacatttgggacaaacctgtgacagccatgccaagagag 2280
 Db 2221 tctcaagctgtgctcactcagacacacatttgggacaaacctgtgacagccatgccaagagag 2280
 QY 2281 gcccttgaccggcccccagagctaaaagcaccagagaaaaatcaatgtcttctactcagcg 2340
 Db 2281 gcccttgaccggcccccagagctaaaagcaccagagaaaaatcaatgtcttctactcagcg 2340
 QY 2341 tgaccacaacttctctagtggtgcaagcgcccccacacctctcgtcagtaaccacacacac 2400
 Db 2341 tgaccacaacttctctagtggtgcaagcgcccccacacctctcgtcagtaaccacacacac 2400
 QY 2401 cactgcttctctcctccacaagtgatctgtatctctagtttcaatttcttttgattga 2460
 Db 2401 cactgcttctctcctccacaagtgatctgtatctctagtttcaatttcttttgattga 2460
 QY 2461 tatgacactataataaatttcatatttgagaatttctcaattgttatctagttaaatagcac 2520
 Db 2461 tatgacactataataaatttcatatttgagaatttctcaattgttatctagttaaatagcac 2520
 QY 2521 agtttgaaactgtctctgagactgacttatacaataatctaacggacaaagatcatatcc 2580
 Db 2521 agtttgaaactgtctctgagactgacttatacaataatctaacggacaaagatcatatcc 2580
 QY 2581 atgtgtatgtgttagacattttatttcatgtactaacccaggaagatctcagtgatgc 2640
 Db 2581 atgtgtatgtgttagacattttatttcatgtactaacccaggaagatctcagtgatgc 2640
 QY 2641 aaatgtgtgcccctgtgtcagctgaaacagctcctgacttcaaaaaacctgaaataag 2700
 Db 2641 aaatgtgtgcccctgtgtcagctgaaacagctcctgacttcaaaaaacctgaaataag 2700
 QY 2701 tctccacagtggtataaatgtgacaatttaggaattttaaactttagatgatttg 2760
 Db 2701 tctccacagtggtataaatgtgacaatttaggaattttaaactttagatgatttg 2760
 QY 2761 ttcaattttatttcaattttattttgttaatgcaaacaggacttaaatgaactttgat 2820
 Db 2761 ttcaattttatttcaattttattttgttaatgcaaacaggacttaaatgaactttgat 2820

Db 2761 ttcaattttatttcaattttattttgttaatgcaaacaggacttaaatgaactttgat 2820
 QY 2821 ctctgttttaaaagattatlaaaaaacattgtgtatctatacatatggtccttgagactt 2880
 Db 2821 ctctgttttaaaagattatlaaaaaacattgtgtatctatacatatggtccttgagactt 2880
 QY 2881 agcttccactacactacagagatgatctccatgtagtccatataaaacctgcagagtga 2940
 Db 2881 agcttccactacactacagagatgatctccatgtagtccatataaaacctgcagagtga 2940
 QY 2941 ttccagaagtgtcagatactgttaattacatctccatgaaggctgaaagaatgacctac 3000
 Db 2941 ttccagaagtgtcagatactgttaattacatctccatgaaggctgaaagaatgacctac 3000
 QY 3001 gttctgtatacagctgt 3060
 Db 3001 gttctgtatacagctgt 3060
 QY 3061 ctgaagctctgcgt 3120
 Db 3061 ctgaagctctgcgt 3120
 QY 3121 tccattcattgtttacgctgtgaaattttctcccatgtgaatgtlaagtaaaacttaagt 3180
 Db 3121 tccattcattgtttacgctgtgaaattttctcccatgtgaatgtlaagtaaaacttaagt 3180
 QY 3181 ttgtcatcaataaatgtgtaataactaaaaaa 3217
 Db 3181 ttgtcatcaataaatgtgtaataactaaaaaa 3217

RESULT 2
 X90786
 ID X90786 standard; DNA; 3190 BP.
 XX
 AC X90786;
 XX
 DT 13-JAN-2000 (first entry)
 XX
 DE DNA encoding human stauflen protein.
 XX
 KW Human stauflen protein; hstau; protein lysate; lung; kidney; testis;
 KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;
 KW telomerase assembly; maturation; transport; regulation; interaction;
 KW inhibitory agent; antibody; immunoassay; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT CDS 345..1835
 FT /*tag= a
 FT /product= "Human stauflen protein"
 FT /note= "Binds to telomerase RNA"
 FT polyA_signal 3172..3177
 FT /*tag= b
 XX
 PN W09951255-A1.
 PN
 PD 14-OCT-1999.
 XX
 PF 06-APR-1999; 99WO-US07533.
 PF
 PR 06-APR-1998; 98US-0080783.
 PR
 XX (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 PA
 PI Greider CW, Le S;
 XX
 DR WPI; 1999-620168/53.
 DR
 DR P-PSDB; Y26335.
 XX
 PT Human stauflen polypeptide useful in methods for identifying telomerase
 PT inhibitors -

Db 1898 ttgaaatttggaatttctgtacactccagctggtccgagagacacgctgtggtlaagat 1957
QY 1976 gtggcgacgacgacgaggaagacacagaacacacgaagcggtgtggtcc-ggctygact 2034
Db 1958 gtggcgacgacgacgaggaagacacagaacacacgaagcggtgtggtccggtgact 2017
QY 2035 gtgctgggtttgtgtgatggtccactcggtgacacctggcggtccctacgcaatagcact 2094
Db 2018 gtgctgggtttgtgtgatggtccactcggtgacacctggcggtccctacgcaatagcact 2077
QY 2095 gcctgtgggaagaagaggtgtgccagccagctgttctccgggacacacgacagatccac 2154
Db 2078 gcctgtgggaagaagaggtgtgccagccagctgttctccgggacacacgacagatccac 2137
QY 2155 accctgggacactccgtgtgtgtcttttttccctgtgtgaaagaagaacgggcacg 2214
Db 2138 accctgggacactccgtgtgtgtcttttttccctgtgtgaaagaagaacgggcacg 2197
QY 2215 acccttccaaagtgtgtctactcagacacacattggtgacaaacctgtgacagccatgccag 2274
Db 2198 acccttccaaagtgtgtctactcagacacacattggtgacaaacctgtgacagccatgccag 2257
QY 2275 agagagcctttagaccggtcccaagctaaagaacacagagaataatcaatgtctctac 2334
Db 2258 agagagcctttagaccggtcccaagctaaagaacacagagaataatcaatgtctctac 2317
QY 2335 tcagcgtgacccaacttctctagtgtgccaggtcccaacctctctgtagtaccacac 2394
Db 2318 tcagcgtgacccaacttctctagtgtgccaggtcccaacctctctgtagtaccacac 2377
QY 2395 catcacactgttctctctccaaagtgtgtatcttagttcatatttcttt 2454
Db 2378 catcacactgttctctctccaaagtgtgtatcttagttcatatttcttt 2437
QY 2455 gattgatgacactatataaaatttcatatttgagaatttccaatgttatctagttaa 2514
Db 2438 gattgatgacactatataaaatttcatatttgagaatttccaatgttatctagttaa 2497
QY 2515 tagcacagtttgaaactgttctgtagactgtattatcaataatctaacggacaaagatc 2574
Db 2498 tagcacagtttgaaactgttctgtagactgtattatcaataatctaacggacaaagatc 2557
QY 2575 atatcatgtgtatgtgttagacattttattcatgtactaacccagagacgtttcag 2634
Db 2558 atatcatgtgtatgtgttagacattttattcatgtactaacccagagacgtttcag 2617
QY 2635 tgatgcaaatgtgtccctctgtgtcagctgaaacagtcctgtgacttcaaaaacctg 2694
Db 2618 tgatgcaaatgtgtccctctgtgtcagctgaaacagtcctgtgacttcaaaaacctg 2677
QY 2695 aatagtctccacagttgtataaattgacaatttagaattttaaacttagatgac 2754
Db 2678 aatagtctccacagttgtataaattgacaatttagaattttaaacttagatgac 2737
QY 2755 attgttccattttatttcatattttattgttaatgcaaacaggacttaaatgac 2814
Db 2738 attgttccattttatttcatattttattgttaatgcaaacaggacttaaatgac 2797
QY 2815 ttgatctctgttttaagaattataaaaaaacattgtgtatctatacatatggtctga 2874
Db 2798 ttgatctctgttttaagaattataaaaaaacattgtgtatctatacatatggtctga 2857
QY 2875 ggaactagcttcaactacactacagatatgctccatgtgtagtccatataaactgcag 2934
Db 2858 ggaactagcttcaactacactacagatatgctccatgtgtagtccatataaactgcag 2917
QY 2935 agtatttccagagtgctcgatactgttaattacatctccattagggtgaaagaatg 2994
Db 2918 agtatttccagagtgctcgatactgttaattacatctccattagggtgaaagaatg 2977
QY 2995 acctagcttctgtatacagctgtgtgtcttttgatgtgtgttactgttacacagaagt 3054

Db 2978 acctagcttctgtatacagctgtgtgtcttttgatgtgtgttactgttacacagaagt 3037
QY 3055 tgtcacctgagctctgcgtgtgtgtccgtatggaacctgtgtagccctgcgagttaagt 3114
Db 3038 tgtcacctgagctctgcgtgtgtgtccgtatggaacctgtgtagccctgcgagttaagt 3097
QY 3115 actgcttccattcatgttttaacgctggaattttctcccatggaatgtgaagtaaaact 3174
Db 3098 actgcttccattcatgttttaacgctggaattttctcccatggaatgtgaagtaaaact 3157
QY 3175 aagtgttgcatacataaatgtgtaataactaa 3207
Db 3158 aagtgttgcatacataaatgtgtaataactaa 3190
RESULT 3
Z93324
ID Z93324 standard; cDNA; 3142 BP.
XX
AC Z93324;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human staufen cDNA.
XX
KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 288..1193
FT /*tag= a
FT /product= Staufen protein
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L;
PI Moulard AJ;
XX
DR WPI: 2000-246924/22.
DR P-PSDB; Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1b; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSF (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This

CC sequence is an alternatively spliced sequence to the one given in
CC Z93322.
XX
SQ Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 94.9%; Score 3053.8; DB 21; Length 3142;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 3140; Conservative 0; Mismatches 2; Indels 75; Gaps 1;

QY 1 acttcctgcgagctgcgggagcgccctgagcgctcttcagcgcttttgcgcgcgctgcgct 60
|||
Db 1 acttcctgcgagctgcgggagcgccctgagcgctcttcagcgcttttgcgcgcgctgcgct 60
QY 61 ctctctcgctccgcctcttcctttgacgcgcctcccccccgcgcccgcgcgccgcctc 120
|||
Db 61 ctctctcgctccgcctcttcctttgacgcgcctcccccccgcgcccgcgcgccgcctc 120
QY 121 ctccacgycgaactccgcctcttcctccctcgctccctctctcctctcccttttccct 180
|||
Db 121 ctccacgycgaactccgcctcttcctccctcgctccctctctcctctcccttttccct 180
QY 181 ctctcctccctcctcgccgcgaaccgcccagcgccgcgcggggagcgagctcgagc 240
|||
Db 181 ctctcctccctcctcgccgcgaaccgcccagcgccgcgcggggagcgagctccgagc 240
QY 241 agcagccagagtttatcaaccacttaacctctcagaactgaacaaagaacaacattgtcc 300
|||
Db 241 agcagcc----- 247
QY 301 tggaaagccctctttttaaanaagaaagcatlaaccctactgttagaactaaatgcactgt 360
|||
Db 248 -----agaaagcataaccctactgttagaactaaatgcactgt 285
QY 361 gcatgaaacttggaaaaaaaccaaattgtataagcctgtttgacccttactctcgatgcagt 420
|||
Db 286 gcatgaaacttggaaaaaaaccaaattgtataagcctgtttgacccttactctcgatgcagt 345
QY 421 ccaacctataactacaacatgagagagtgctatacccccgaggtactttaccatttc 480
|||
Db 346 ccaacctataactacaacatgagagagtgctatacccccgaggtactttaccatttc 405
QY 481 cagttccaaccttactatatacaagtgaactctctgttggagagcagacaatttaatgca 540
|||
Db 406 cagttccaaccttactatatacaagtgaactctctgttggagagcagacaatttaatgca 465
QY 541 aaggaagaacagacagcgctgcgaacacgattgctgcgaagcggttgaggtacctgc 600
|||
Db 466 aaggaagaacagacagcgctgcgaacacgattgctgcgaagcggttgaggtacctgc 525
QY 601 agaattgagccctgcgagagagcgctgaggttgaattggaagaaatccgaagaagaaatc 660
|||
Db 526 agaattgagccctgcgagagagcgctgaggttgaattggaagaaatccgaagaagaaatc 585
QY 661 tcaataaacttgaataaagtcaagtgtttgagattgcacttaacggaacttgctgtga 720
|||
Db 586 tcaataaacttgaataaagtcaagtgtttgagattgcacttaacggaacttgctgtga 645
QY 721 atttcgagtggtgcccgggagagtggtgccaccaccacatgaagaactttgtgccaagt 780
|||
Db 646 atttcgagtggtgcccgggagagtggtgccaccaccacatgaagaactttgtgccaagt 705
QY 781 cggttggggagtttggggggaaggtgaagggaanaagcaagaagatttcaaagaaatg 840
|||
Db 706 cggttggggagtttggggggaaggtgaagggaanaagcaagaagatttcaaagaaatg 765
QY 841 ccgcatagctgttcttgagagagctgaagaagttaaccgcccctgcctgcagttgaacgag 900
|||
Db 766 ccgcatagctgttcttgagagagctgaagaagttaaccgcccctgcctgcagttgaacgag 825
QY 901 taaagcctagaattcaaaaagaaaaacaacaccatagtcaagccacagacaagccagaat 960
|||
Db 826 taaagcctagaattcaaaaagaaaaacaacaccatagtcaagccacagacaagccagaat 885

QY 961 atggccaggggatcaatcccgatttagccgactgcccagatccagcagcnaaaaaagaga 1020
|||
Db 886 atggccaggggatcaatcccgatttagccgactgcccagatccagcagcnaaaaaagaga 945
QY 1021 aggaagcagagtagaacgtctctcacagagcgagcgctccgcgcgcgaggagtttgtga 1080
|||
Db 946 aggaagcagagtagaacgtctctcacagagcgagcgctccgcgcgcgaggagtttgtga 1005
QY 1081 tgcaggtgaaggttggaaaaccacacttgcagaaggaacggtgcaaccaagaaggtggcca 1140
|||
Db 1006 tgcaggtgaaggttggaaaaccacacttgcagaaggaacggtgcaaccaagaaggtggcca 1065
QY 1141 agcgcaatgcagccgagaaacatgtgtgagatccttggtttcaaaagtccgcgagcgagc 1200
|||
Db 1066 agcgcaatgcagccgagaaacatgtgtgagatccttggtttcaaaagtccgcgagcgagc 1125
QY 1201 ccaaccaaacccgactcaagtcaaggagaaagacccaataaagaacagggtgaa 1260
|||
Db 1126 ccaaccaaacccgactcaagtcaaggagaaagacccaataaagaacagggtgaa 1185
QY 1261 gaaaagtaacctttttgaaacctgtgcttgggatagaaatggtgactagtaataagag 1320
|||
Db 1186 gaaaagtaacctttttgaaacctgtgcttgggatagaaatggtgactagtaataagag 1245
QY 1321 atgagttcaggaatgccttaataagtcatcagcagctgcctgtgtgaattcttccatg 1380
|||
Db 1246 atgagttcaggaatgccttaataagtcatcagcagctgcctgtgtgaattcttccatg 1305
QY 1381 tgcgcgaggtcgcccgagctgtlagagttagtcagaagacatcacaccaagaatttacc 1440
|||
Db 1306 tgcgcgaggtcgcccgagctgtlagagttagtcagaagacatcacaccaagaatttacc 1365
QY 1441 gggcagctccgaatccttgcgaagggccaaggttaactgccaatgatagccgagagttgt 1500
|||
Db 1366 gggcagctccgaatccttgcgaagggccaaggttaactgccaatgatagccgagagttgt 1425
QY 1501 atgggggcacctgcgcccaagcccgagaccattttaagaataacatctctcaggccag 1560
|||
Db 1426 atgggggcacctgcgcccaagcccgagaccattttaagaataacatctctcaggccag 1485
QY 1561 taaccatggaacctctcaagagaccctctgagcaactggaactatcttccagagtcag 1620
|||
Db 1486 taaccatggaacctctcaagagaccctctgagcaactggaactatcttccagagtcag 1545
QY 1621 gatccaggttgaatacaaaagacttccccaaaaacaagaagaagaaattgtatctcta 1680
|||
Db 1546 gatccaggttgaatacaaaagacttccccaaaaacaagaagaagaaattgtatctcta 1605
QY 1681 tcaattgtcctctcagccacctctgatacagccaatgtalatcggaagagttgagtcct 1740
|||
Db 1606 tcaattgtcctctcagccacctctgatacagccaatgtalatcggaagagttgagtcct 1665
QY 1741 gccatgatatgtgtgctgaacatcttaagtgtgtctgagtggaccaacaagta 1800
|||
Db 1666 gccatgatatgtgtgctgaacatcttaagtgtgtctgagtggaccaacaagta 1725
QY 1801 cagagatgccaaagaacaggaaacggaccaatgtctgtgttgggaggtgtgaaccttt 1860
|||
Db 1726 cagagatgccaaagaacaggaaacggaccaatgtctgtgttgggaggtgtgaaccttt 1785
QY 1861 ctggccaatgaaccattataaaaatcccaacatatactgaaaaatactgaaactgtttga 1920
|||
Db 1786 ctggccaatgaaccattataaaaatcccaacatatactgaaaaatactgaaactgtttga 1845
QY 1921 aaatttgaatttctgatacctccagttggccgagagacaacggttggtaaaagatgtgg 1980
|||
Db 1846 aaatttgaatttctgatacctccagttggccgagagacaacggttggtaaaagatgtgg 1905
QY 1981 cagcagcaggggaagacacagaaacacaacagagggcggtgtgtgcccgtcgactgtgt 2040
|||
Db 1906 cagcagcaggggaagacacagaaacacaacagagggcggtgtgtgcccgtcgactgtgt 1965

QY	2041	gggttgtgtgtatgagccactcgtgtgacctggtccctacgcaatagcagctgctgt	2100
Db	1966	gggttgtgtgtatgagccactcgtgtgacctggtccctacgcaatagcagctgctgt	2025
QY	2101	ggggaagaagggtgccccagccagctgttctcccgggacaccagcagatccacacctg	2160
Db	2026	ggggaagaagggtgccccagccagcagctgttctcccgggacaccagcagatccacacctg	2085
QY	2161	ggcacctccgtgttgtgtcttttttccctgtgtgaagaagaacgycagcacccct	2220
Db	2086	ggcacctccgtgttgtgtcttttttccctgtgtgaagaagaacgycagcacccct	2145
QY	2221	tctcaagctgtgctcaactcagacacacatltgaggacaaacccctygacagccatgccaagagag	2280
Db	2146	tctcaagctgtgctcaactcagacacacatltgaggacaaacccctygacagccatgccaagagag	2205
QY	2281	gcctltgaccggccccagagctaaaagcacccagagaaaatcaaatgcttccctactcagcg	2340
Db	2206	gcctltgaccggccccagagctaaaagcacccagagaaaatcaaatgcttccctactcagcg	2265
QY	2341	tgaaccaactttcttagtgtgtgccacggccccaccacactcctctgcagtlaccacaccatcac	2400
Db	2266	tgaaccaactttcttagtgtgtgccacggccccaccacactcctctgcagtlaccacaccatcac	2325
QY	2401	cactgcttctcttcccaacagtgatctgtatctttagtlttcaatlatttctttgtatga	2460
Db	2326	cactgcttctcttcccaacagtgatctgtatctttagtlttcaatlatttctttgtatga	2385
QY	2461	tatgacactataataaattttcatlttgagaatttctcaatgtlatctagtttaaatagcac	2520
Db	2386	tatgacactataataaattttcatlttgagaatttctcaatgtlatctagtttaaatagcac	2445
QY	2521	agtttggaaactgtctgagactgactttatccaataatctaacccgacaaagatcatatcc	2580
Db	2446	agtttggaaactgtctgagactgactttatccaataatctaacccgacaaagatcatatcc	2505
QY	2581	atgtgtatgtgttagaacattttatltcatltgactaaaccagagacagtttcagtgtatgc	2640
Db	2506	atgtgtatgtgttagaacattttatltcatltgactaaaccagagacagtttcagtgtatgc	2565
QY	2641	aaatgtgtgcccctcgtgttcagctgnaaacagtlccctggaactltcaaaaaccttgataag	2700
Db	2566	aaatgtgtgcccctcgtgttcagctgnaaacagtlccctggaactltcaaaaaccttgataag	2625
QY	2701	tctcccaacagttgtataaatttgacaatttaggaattttaaactttagatgatcatlttg	2760
Db	2626	tctcccaacagttgtataaatttgacaatttaggaattttaaactttagatgatcatlttg	2685
QY	2761	ttccatttttatttcaatttttattttgttlaatgcaaaacaggaacttaaatgaaactttgat	2820
Db	2686	ttccatttttatttcaatttttattttgttlaatgcaaaacaggaacttaaatgaaactttgat	2745
QY	2821	ctctgttttaagaatatttaaaaaacattgtgtatctatcatatgtgctcttgaggactt	2880
Db	2746	ctctgttttaagaatatttaaaaaacattgtgtatctatcatatgtgctcttgaggactt	2805
QY	2881	agccttcaactacactacaggaatatgatctccatgtagtccaataaaacctgcagagtgtat	2940
Db	2806	agccttcaactacactacaggaatatgatctccatgtagtccaataaaacctgcagagtgtat	2865
QY	2941	tttccagagtgctcgataactgttlaattacatctccattagggctgnaaaagaaatgacctac	3000
Db	2866	tttccagagtgctcgataactgttlaattacatctccattagggctgnaaaagaaatgacctac	2925
QY	3001	gtttctgtatacagctgtgtgtgtcttttgatgtgtgtgttactgttacacagaagtggtgca	3060
Db	2926	gtttctgtatacagctgtgtgtgtcttttgatgtgtgtgttactgttacacagaagtggtgca	2985
QY	3061	ctgaagctctgcgtgtgtgtccgtatgnaaaacctgtgtagccctgcgagttaaagtactgct	3120
Db	2986	ctgaagctctgcgtgtgtgtccgtatgnaaaacctgtgtagccctgcgagttaaagtactgct	3045
QY	3121	tcacattcaattgttlaacgctgygaatttttctcccccagtgaaatgtgaagtaaaacttaagtgt	3180

Db 3046 tccattcatgtgttaacgtggaattttctcccatggaatgtaagtaaacttaagTgt 3105

QY 3181 ttgtcatcaataaatggtataactataaaaaaaaaa 3217

Db 3106 ttgtcatcaataaatggtataactataaaaaaaaaa 3142

RESULT	4	
Z93327		
ID	Z93327	standard; cDNA; 3260 BP.
XX		
AC	Z93327;	
XX		
DT	04-JUL-2000	(first entry)
XX		
DE	Human staufen cDNA.	
XX		
KW	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW	acquired immune deficiency syndrome; protease; ; human; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	409..1311
FT		/*tag= a
FT		/product= Staufen protein

PN	CA2238656-A1.
XX	
PD	22-NOV-1999.
XX	
PF	22-MAY-1998; 98CA-2238656.
XX	
PR	22-MAY-1998; 98CA-2238656.
XX	
PA	(UYMO-) UNIV MONTREAL.
XX	
PI	Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI	Mouland AJ;
XX	
WPI;	2000-246924/22.
DR	P-PSDB; Y83108.
DR	

PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)

PS Claim 4; Figure 1b; 96pp; English.

Staufen is a RNA binding protein which interacts with double stranded RNA and/or rough endoplasmic reticulum. It is involved in mRNA transport and localization. Mammalian SP contain multiple ds RNA-binding domains. Chimeric proteins comprising portions of staufen with proteins with RNase or protease activity can be used to treat virus infections. The RNase or protease activity of the fusion protein prevents proper maturation of the virus. Mammalian staufen proteins seem to recognise double stranded RNA structure rather than any sequence specific position. 2 bacteriially-expressed fusion proteins used in an RNA-binding assay (his/hSP (human SP) and MBP/mSP (murine SP)) both proteins strongly bound double stranded RNA. Both fusion proteins also directly bound labeled ds RNAs and RNA/DNA hybrids. Staufen proteins may therefore be useful in the treatment of RNA virus infections, especially those caused by retroviruses, in particular human immunodeficiency virus. This sequence is an alternatively spliced sequence to the one given in Z93332.

Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

Query Match	94.48;	Score 3037.4;	DB 21;	Length 3260;
-------------	--------	---------------	--------	--------------

Best Local Similarity 96.6%; Pred. No. 0;
Matches 3153; Conservative 0; Mismatches 61; Indels 49; Gaps 3;

QY 1 actctcgcgcggtctgcggcgccctgagcgcctctcagcgcttctgcgcgcgctgcgcgt 60
Db 1 actctcgcgcggtctgcggcgccctgagcgcctctcagcgcttctgcgcgcgctgcgcgt 60
QY 61 ctctctcgcgtcccgcttccctttgacgcgcctcccccccgccgcgcgcgcgcgcctc 120
Db 61 ctctctcgcgtcccgcttccctttgacgcgcctcccccccgccgcgcgcgcgcgcctc 120
QY 121 ctccacgcgcactccgcctcttccctcccttcgctcc----- 157
Db 121 ctccacgcgcactccgcctcttccctcccttcagcagcagccagcagcgagtgagtg 180
QY 158 -ttcttcctctcccttttctctcttctccttccctcctcgcgcgcacgcgcagcg 216
Db 181 ttctcctctgtcgcgccagcgtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 240
QY 217 ccgcgcgcgggga-----cgagctcgagcagcagcgagtggtt 254
Db 241 cctccacggtcagcgatcttcccaacttcagcctcccgataagctgagatcacaagagttt 300
QY 255 attaaccaacttaacctctcagaactgaaacaaagacaacatgttctctgaaacgcctct 314
Db 301 attaaccaacttaacctctcagaactgaaacaaagacaacatgttctctgaaacgcctct 360
QY 315 tttaaaaaaagaagcataaccctactctgtagaactaaatgcactgtgcataaacttga 374
Db 361 tttaaaaaaagaagcataaccctactctgtagaactaaatgcactgtgcataaacttga 420
QY 375 aaaaaaaccaatgtataagcctgttgaccctactcctcgcgcgcacgcacataactac 434
Db 421 aaaaaaaccaatgtataagcctgttgaccctactcctcgcgcgcacgcacataactac 477
QY 435 aacatgagagagtggtctatcccccgaggtactttaccatltccagttccaccttta 494
Db 478 aacatgagagagtggtctatcccccgaggtactttaccatltccagttccaccttta 537
QY 495 cttatcaagtgaacttctgttgaggagcagcgaatttaatgccaaggaagacaaga 554
Db 538 cttatcaagtgaacttctgttgaggagcagcgaatttaatgccaaggaagacaaga 597
QY 555 caggtctcgaaaacagatgctgtcgcaaaagcgttgaggtcctgcagaaatgagccctg 614
Db 598 caggtctcgaaaacagatgctgtcgcaaaagcgttgaggtcctgcagaaatgagccctg 657
QY 615 ccaagagaggtcgtgaagtgaatggaagaaatccgaaagaagaaatctcaataaatctgaa 674
Db 658 ccaagagaggtcgtgaagtgaatggaagaaatccgaaagaagaaatctcaataaatctgaa 717
QY 675 ataagtcgaagtgtttgagattgcaacttaaacgcgaacttgctgtgaatttcgaggtgccc 734
Db 718 ataagtcgaagtgtttgagattgcaacttaaacgcgaacttgctgtgaatttcgaggtgccc 777
QY 735 cgggagagtggtcccaaccacatgaagaacttgtgaccaaggttcggttggggagttt 794
Db 778 cgggagagtggtcccaaccacatgaagaacttgtgaccaaggttcggttggggagttt 837
QY 795 gtgggggaaggtgaagggaagaaagcaagaagatttcaagaagaaatgcgcgcacatagctgtt 854
Db 838 gtgggggaaggtgaagggaagaaagcaagaagatttcaagaagaaatgcgcgcacatagctgtt 897
QY 855 cttaggagagctgaagaagttaacgccccctgcctgcagttgaacgagtaaaagcctagaatc 914
Db 898 cttaggagagctgaagaagttaacgccccctgcctgcagttgaacgagtaaaagcctagaatc 957
QY 915 aaaaaaacaacaacaacatagtcgaagccacagacaagccagatatgtgcaggggagtc 974
Db 958 aaaaaaacaacaacaacatagtcgaagccacagacaagccagatatgtgcaggggagtc 1017
QY 975 aatccgattagcggaactgtgccagatccagcaggcacaaaaggaagaggagccagagttac 1034
|||||

Db 1018 aatccgattagcggaactgtgccagatccagcaggcacaaaaggaagaggccaaggtac 1077
QY 1035 acgctcctcacagagcgaggtcctcccgcgccgcgcagggagtttgtatgcaggtgaaggtt 1094
Db 1078 acgctcctcacagagcgaggtcctcccgcgccgcgcagggagtttgtatgcaggtgaaggtt 1137
QY 1095 ggaaccacactgcagaaggaacggtgcaccaacaagaaggttgcgaagcgcaatgcagcc 1154
Db 1138 ggaaccacactgcagaaggaacggtgcaccaacaagaaggttgcgaagcgcaatgcagcc 1197
QY 1155 gagaacatgtcgtgagatccttggtttcaaaagtcctcgcaagcgagcccaaccaaccgca 1214
Db 1198 gagaacatgtcgtgagatccttggtttcaaaagtcctcgcaagcgagcccaaccaaccgca 1257
QY 1215 ctcaagtcagaggaagacaaccaataaagaacaggggagtgaagaagaaagtaacctt 1274
Db 1258 ctcaagtcagaggaagacaaccaataaagaacaggggagtgaagaagaaagtaacctt 1317
QY 1275 ttgaaacctggtctggtggtgtaaaatggtgactagtaataaagagatgagttcagatg 1334
Db 1318 ttgaaacctggtctggtggtgtaaaatggtgactagtaataaagagatgagttcagatg 1377
QY 1335 cctatctaagtcacagcagctgcgtctgtgaattcttccatgtgtcccgagtgccc 1394
Db 1378 cctatctaagtcacagcagctgcgtctgtgaattcttccatgtgtcccgagtgccc 1437
QY 1395 caggtctgtagagttagtcaaaggacatcacaccaagaatttaccagggcagctccgaat 1454
Db 1438 caggtctgtagagttagtcaaaggacatcacaccaagaatttaccagggcagctccgaat 1497
QY 1455 cctgtccaaggccacggttaactgcacatgatagcccgagagttgttgtatggtggcacctcg 1514
Db 1498 cctgtccaaggccacggttaactgcacatgatagcccgagagttgttgtatggtggcacctcg 1557
QY 1515 cccaagcgcgagacatlttaagaataacatctcttcaggccaagtaaccacatgagcct 1574
Db 1558 cccaagcgcgagacatlttaagaataacatctcttcaggccaagtaaccacatgagcct 1617
QY 1575 ctcaagagaccctctgaagcaactggaactatcttccagagttccagggattccagttgaa 1634
Db 1618 ctcaagagaccctctgaagcaactggaactatcttccagagttccagggattccagttgaa 1677
QY 1635 tacaagacttccccaaaacaacaagaagaaattgtatctcttaaccaattgtcctct 1694
Db 1678 tacaagacttccccaaaacaacaagaagaaattgtatctcttaaccaattgtcctct 1737
QY 1695 cagccaacctgtatcagccatgtgtatcggcaaggatgtgagttcctgcacatgatgtgct 1754
Db 1738 cagccaacctgtatcagccatgtgtatcggcaaggatgtgagttcctgcacatgatgtgct 1797
QY 1755 gcgctgaacatctttaaagttgtctgtctgagttggaaccaacaagaatcacagagatgcacaaga 1814
Db 1798 gcgctgaacatctttaaagttgtctgtctgagttggaaccaacaagaatcacagagatgcacaaga 1857
QY 1815 acaggaacaggaaccaatgtctgtgttggaaggtctgaaaccttcttgccatgaaacca 1874
Db 1858 acaggaacaggaaccaatgtctgtgttggaaggtctgaaaccttcttgccatgaaacca 1917
QY 1875 ttataaatcccaacatatatactgaaataactgaaactgtcttgaanaatttgaatttc 1934
Db 1918 ttataaatcccaacatatatactgaaataactgaaactgtcttgaanaatttgaatttc 1977
QY 1935 tgatacctccagttgtggtcgagagacacggtgtggttaaagatgttgggcagcagcaggaag 1994
Db 1978 tgatacctccagttgtggtcgagagacacggtgtggttaaagatgttgggcagcagcaggaag 2037
QY 1995 acaacagaacacacaggaagcggtgtgtgcccgtgtgactgtgtcgtgggttgtgtgtat 2054
Db 2038 acaacagaacacacaggaagcggtgtgtgcccgtgtgactgtgtcgtgggttgtgtgtat 2097
QY 2055 ggcactcgtgtgacctgtggttccctacgcaatagcagctgtcgttggggaagaggtc 2114
Db 2098 ggcactcgtgtgacctgtggttccctacgcaatagcagctgtcgttggggaagaggtc 2157

QY 2115 gccacgccagctggtctcccgagacaccagcagatccaacccctgggcacctccgtgtt 2174
|||||
Db 2158 gccacgccagctggtctcccgagacaccagcagatccaacccctgggcacctccgtgtt 2217
QY 2175 tggctcttttttccccctgtgtgaagaagaacgcgcagaccccttctcaagctggctc 2234
|||||
Db 2218 tggctcttttttccccctgtgtgaagaagaacgcgcagaccccttctcaagctggctc 2277
QY 2235 actcagacacatttgggacaaaaccttggacagccatgccagagagagccttggaccggcc 2294
|||||
Db 2278 actcagacacatttgggacaaaaccttggacagccatgccagagagagccttggaccggcc 2337
QY 2295 ccagagctaaaagcaccagagaaaaatcaaatgcttcctactcagcgtgacccaactttc 2354
|||||
Db 2338 ccagagctaaaagcaccagagaaaaatcaaatgcttcctactcagcgtgacccaactttc 2397
QY 2355 tagtgtgcccgcccccacacccctcctgcagtaccacacacatcaccaactgtctctct 2414
|||||
Db 2398 tagtgtgcccgcccccacacccctcctgcagtaccacacacatcaccaactgtctctct 2457
QY 2415 ccaacagtgatctgtatcttcttagtttcatttcttcttggattgatatgacactata 2474
|||||
Db 2458 ccaacagtgatctgtatcttcttagtttcatttcttcttggattgatatgacactata 2517
QY 2475 aaatttcaatttgagaatttctcaattgtatctagttaaatagcacagtttggaaacttg 2534
|||||
Db 2518 aaatttcaatttgagaatttctcaattgtatctagttaaatagcacagtttggaaacttg 2577
QY 2535 tctgagactgacttatacaataatcaaacgcagacaaagatcatatccatgtgtatgtgtt 2594
|||||
Db 2578 tctgagactgacttatacaataatcaaacgcagacaaagatcatatccatgtgtatgtgtt 2637
QY 2595 agacattttatttcatcttgactaaccacgagcagtttcagtgatgcaaatgtgtgcct 2654
|||||
Db 2638 agacattttatttcatcttgactaaccacgagcagtttcagtgatgcaaatgtgtgcct 2697
QY 2655 ctggttcagctgaaacagctcctgtgactttcaaaaaccttgaaataagctctccacagttgt 2714
|||||
Db 2698 ctggttcagctgaaacagctcctgtgactttcaaaaaccttgaaataagctctccacagttgt 2757
QY 2715 ataattgacaatttaggaatttaaacctttagatgatcatttggttccattttattt 2774
|||||
Db 2758 ataattgacaatttaggaatttaaacctttagatgatcatttggttccattttattt 2817
QY 2775 cattttatttggtttaatgcaaacaggaacttaaatgaacttgatctctgttttaaga 2834
|||||
Db 2818 cattttatttggtttaatgcaaacaggaacttaaatgaacttgatctctgttttaaga 2877
QY 2835 ttattaaaaacatgtgtatctatatacatatgctcttgaggacttagctttcacatac 2894
|||||
Db 2878 ttattaaaaacatgtgtatctatatacatatgctcttgaggacttagctttcacatac 2937
QY 2895 tacagatatgatctccatgtagtccatataaaacctgcagagtgattttccagagtgctc 2954
|||||
Db 2938 tacagatatgatctccatgtagtccatataaaacctgcagagtgattttccagagtgctc 2997
QY 2955 gatactgttaattacatctccatagggtgaaagaatgacctaacglttctgtatacag 3014
|||||
Db 2998 gatactgttaattacatctccatagggtgaaagaatgacctaacglttctgtatacag 3057
QY 3015 ctgtgttcttttgatgtgtgttactgtacacagaagtgtgcaactgagcgtcgtcgt 3074
|||||
Db 3058 ctgtgttcttttgatgtgtgttactgtacacagaagtgtgcaactgagcgtcgtcgt 3117
QY 3075 gtgtccgtatggaaaacctgtgtagccctgcgagttaagtactgcttccattcatgttt 3134
|||||
Db 3118 gtgtccgtatggaaaacctgtgtagccctgcgagttaagtactgcttccattcatgttt 3177
QY 3135 acgctggaattttctcccatgagatgttaagtaaaacttaagtgtttgtcatcaataaa 3194
|||||
Db 3178 acgctggaattttctcccatgagatgttaagtaaaacttaagtgtttgtcatcaataaa 3237

QY 3195 tggtaataactaaaaaaa 3217
|||||
Db 3238 tggtaataactaaaaaaa 3260
RESULT 5
293322
ID 293322 standard; cDNA; 3506 BP.
XX
AC 293322;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 409..2142
FT /tag= a
FT /product= Stauflen
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PE 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;
PI Moulond AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83023.
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96pp; English.
XX
CC Stauflen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC stauflen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian stauflen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hSP (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauflen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

Query Match 90.6%; Score 2914.8; DB 21; Length 3506;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 3215; Conservative 0; Mismatches 2; Indels 289; Gaps 1;

QY 1 acttcctgcgggctgcggcgctgtgagcgtcttcagcgtttgcggcgcgctgcgcgt 60
|||||

Db 1 acttcctgcgpgctgcgpgcgccctgagcgctcttcagcgcttctgcgpgcgctgcgcgt 60
QY 61 ctctctcggtcccgctctcctttagacgcctcccccccgcgcccgcgccgcgcctc 120
Db 61 ctctctcggtcccgctctcctttagacgcctcccccccgcgcccgcgccgcgcctc 120
QY 121 ctccacggcactccgcctcttcctcccttcgctccctcttcctctcccttttccct 180
Db 121 ctccacggcactccgcctcttcctcccttcgctccctcttcctctcccttttccct 180
QY 181 ctccctccccctcctgcgcgcaaccgcccagagccgcgpggagcgagctcgagc 240
Db 181 ctccctccccctcctgcgcgcaaccgcccagagccgcgpggagcgagctccgagc 240
QY 241 agcagccagagtttattaaccacttaacctctcagaaactgaaacaacacattgttc 300
Db 241 agcagccagagtttattaaccacttaacctctcagaaactgaaacaacacattgttc 300
QY 301 tggaaagccctctttttaaaa----- 322
Db 301 tggaaagccctctttttaaaaagtagaactttagacttcatagcactgaaattaaactg 360
QY 323 ----- 322
Db 361 cactgaaagctgtttacctgcaltgttcactttgtgaaagtgcacatgctcaagtt 420
QY 323 ----- 322
Db 421 caagtgcaagttcagaaccocatctgctctctcagggagccaataactgaaacaagac 480
QY 323 ----- 322
Db 481 cagtcctctctctcacagcctttagtgagtatcccttctactactagctctctgcctct 540
QY 323 ----- 322
Db 541 gaaaatgcaggtagaccocatcaaaactcgtcttaaccctctgcatactataatccacc 600
QY 323 -----agaaagcataaccctactgtgtagaactaaatgcactgtgcatagaaact 371
Db 601 agtgcagctgcagaaagcataaccctactgttagaactaaatgcactgtgcatagaaact 660
QY 372 ggaaaaaaaaccaatgtataagcctgttgacccttactctcgaatgcagtcacactataac 431
Db 661 ggaaaaaaaaccaatgtataagcctgttgacccttactctcgaatgcagtcacactataac 720
QY 432 taacaatgagaggtgtctatcccccgaggtactttaccatttccagttccacct 491
Db 721 taacaatgagaggtgtctatcccccgaggtactttaccatttccagttccacct 780
QY 492 ttactttatcaagtgcgaactttctgttgaggagagcaagcaatttaatgcaaaaggaaagaca 551
Db 781 ttactttatcaagtgcgaactttctgttgaggagagcaagcaatttaatgcaaaaggaaagaca 840
QY 552 agacaggtcgcgaacaacgaatgctgcgcgaagcgttgagatccctgcagaatlgagccc 611
Db 841 agacaggtcgcgaacaacgaatgctgcgcgaagcgttgagatccctgcagaatlgagccc 900
QY 612 ctgcagagaggtgcgaggtgaatggaagagaatccgaagaagaataatcacaataact 671
Db 901 ctgcagagaggtgcgaggtgaatggaagagaatccgaagaagaataatcacaataact 960
QY 672 gaaataagtcaagtgtttgagatttgcaacttaaacggaacttgctgtgaatttcgagtg 731
Db 961 gaaataagtcaagtgtttgagatttgcaacttaaacggaacttgctgtgaatttcgagtg 1020
QY 732 gcccgaggaggtgcgccacccacatgaaagaacttgtgaccaaagttccggttgaggag 791
Db 1021 gcccgaggaggtgcgccacccacatgaaagaacttgtgaccaaagttccggttgaggag 1080
QY 792 ttctgtgggggaaagtgaagggaaaaagcaagaagattcaagaaaaatgcgcgcatagct 851
Db 1081 ttctgtgggggaaagtgaagggaaaaagcaagaagattcaagaaaaatgcgcgcatagct 1140

QY 852 gtctctgagagctgaagaagttaccgccccctgcctgcagttgaaagagtaaagcctaga 911
Db 1141 gtctctgagagctgaagaagttaccgccccctgcctgcagttgaaagagtaaagcctaga 1200
QY 912 atcaaaaaagaacaacaacccatagtcgaagccacagacaagccagatatgcccagggg 971
Db 1201 atcaaaaaagaacaacaacccatagtcgaagccacagacaagccagatatgcccagggg 1260
QY 972 atcaatccgattagccgactgcccagatccagcagggcaaaaaaggaagaagccagag 1031
Db 1261 atcaatccgattagccgactgcccagatccagcagggcaaaaaaggaagaagccagag 1320
QY 1032 tacacgctctcacaagagcgagcgctcccgccgcgagggagtttgtgattgcaggtgaag 1091
Db 1321 tacacgctctcacaagagcgagcgctcccgccgcgagggagtttgtgattgcaggtgaag 1380
QY 1092 gtcgaaaccacactgcgaaaggaacgggcacaacaagaaggtggccaagcgcaatgca 1151
Db 1381 gtcgaaaccacactgcgaaaggaacgggcacaacaagaaggtggccaagcgcaatgca 1440
QY 1152 gccgagaacatgctgtagatccttgtttcaagttcccgacgpgcagcccaaaccc 1211
Db 1441 gccgagaacatgctgtagatccttgtttcaagttcccgacgpgcagcccaaaccc 1500
QY 1212 gcactcaagtcagaggaagaagacacccataaagaacccagggagttgaaagaaagtaacc 1271
Db 1501 gcactcaagtcagaggaagaagacacccataaagaacccagggagttgaaagaaagtaacc 1560
QY 1272 tttttgaaacctggtctgggagtgaaaaatgggactagtaataaagagatgagttcagg 1331
Db 1561 tttttgaaacctggtctgggagtgaaaaatgggactagtaataaagagatgagttcagg 1620
QY 1332 atgccttatcctaagtcacagcagctgcctgcgtggaattcttcccatggtgccgagtc 1391
Db 1621 atgccttatcctaagtcacagcagctgcctgcgtggaattcttcccatggtgccgagtc 1680
QY 1392 gcccaagctgtagaggttagtcaagagacatcacacccaagaattttaccagggcagctccg 1451
Db 1681 gcccaagctgtagaggttagtcaagagacatcacacccaagaattttaccagggcagctccg 1740
QY 1452 aatcctgcgaagggccaaggttaactgccaatgataagcccagagttgttgtatggggcacc 1511
Db 1741 aatcctgcgaagggccaaggttaactgccaatgataagcccagagttgttgtatggggcacc 1800
QY 1512 tcgcccacagccgagaccattttaaagaataacatctctcagggccaagtaaccatgga 1571
Db 1801 tcgcccacagccgagaccattttaaagaataacatctctcagggccaagtaaccatgga 1860
QY 1572 cctctcacgagaccctctgagcaactggaactatcttccagagttccaggatccaggtt 1631
Db 1861 cctctcacgagaccctctgagcaactggaactatcttccagagttccaggatccaggtt 1920
QY 1632 gaatacaaaagacttccccaataaacaacaagaacgaatttgtatctctatacaattgctcc 1691
Db 1921 gaatacaaaagacttccccaataaacaacaagaacgaatttgtatctctatacaattgctcc 1980
QY 1692 tctcagccacctctgatacagccatggtatcggaaggaatgttgaggtccctgccaatgatatg 1751
Db 1981 tctcagccacctctgatacagccatggtatcggaaggaatgttgaggtccctgccaatgatatg 2040
QY 1752 gctgcgctgaacatctttaaagttgctgtctgagttggaaccaaaagtaacagatgcca 1811
Db 2041 gctgcgctgaacatctttaaagttgctgtctgagttggaaccaaaagtaacagatgcca 2100
QY 1812 agaacaggaacggaaccaatgtctgtgtgtggaggtgctgaaccttcttgcccatgaa 1871
Db 2101 agaacaggaacggaaccaatgtctgtgtgtggaggtgctgaaccttcttgcccatgaa 2160
QY 1872 ccattataaaatcccaacatatatactgaaaaactgnaaactgcttggaaaatttggaaat 1931
Db 2161 ccattataaaatcccaacatatatactgaaaaactgnaaactgcttggaaaatttggaaat 2220

QY 1932 ttctgatacctccaagtggccgagagacacagtggtgtaaagatgttggcagcagcagg 1991
|||||
Db 2221 ttctgatacctccaagtggccgagagacacagtggtgtaaagatgttggcagcagcagg 2280
QY 1992 aagacaacagaaacacaaaggaggcggtgtgtgcccgttgactgtgtgggtttgtgt 2051
|||||
Db 2281 aagacaacagaaacacaaaggaggcggtgtgtgcccgttgactgtgtgggtttgtgt 2340
QY 2052 gatggccaactcggtgacctggcggtccctacgcgaatagcagctgcctgttgggaaagaag 2111
|||||
Db 2341 gatggccaactcggtgacctggcggtccctacgcgaatagcagctgcctgttgggaaagaag 2400
QY 2112 gctgcccagccaagtgttctctcccgagacaccagagatccacaccttgggcacctcgt 2171
|||||
Db 2401 gctgcccagccaagtgttctctcccgagacaccagagatccacaccttgggcacctcgt 2460
QY 2172 gtttggctcttttttccctgtgtgaaagaagaacgacgacccctctcaagctgg 2231
|||||
Db 2461 gtttggctcttttttccctgtgtgaaagaagaacgacgacccctctcaagctgg 2520
QY 2232 ctcaactcagacacacatltgggacaaaccttgagcagccatgccaagagagagcctttgaccg 2291
|||||
Db 2521 ctcaactcagacacacatltgggacaaaccttgagcagccatgccaagagagagcctttgaccg 2580
QY 2292 gccccagagctaaaaagcacagagaanaatcaaatgtctccctactcagcgtgaccccaactt 2351
|||||
Db 2581 gccccagagctaaaaagcacagagaanaatcaaatgtctccctactcagcgtgaccccaactt 2640
QY 2352 ttctagtgtgcacgcccccaaccctctcgtcagtaaccaacacacatcaaccaactgtcttct 2411
|||||
Db 2641 ttctagtgtgcacgcccccaaccctctcgtcagtaaccaacacacatcaaccaactgtcttct 2700
QY 2412 ctccaacagtgatctgtatcttctagtcttcatatttcttttgattgatgatgacactat 2471
|||||
Db 2701 ctccaacagtgatctgtatcttctagtcttcatatttcttttgattgatgatgacactat 2760
QY 2472 ataaaaatttcaatttgagaatttccaattgtatctagttaaatagcacagcttggaaac 2531
|||||
Db 2761 ataaaaatttcaatttgagaatttccaattgtatctagttaaatagcacagcttggaaac 2820
QY 2532 ttgtctgagactgactttatccaataatctaaccgacaaagatcatatccatgtgtatgtg 2591
|||||
Db 2821 ttgtctgagactgactttatccaataatctaaccgacaaagatcatatccatgtgtatgtg 2880
QY 2592 gttagacatttttatttcatattgactaaaccagagacagtttcagtgatgcaaatgtgtgc 2651
|||||
Db 2881 gttagacatttttatttcatattgactaaaccagagacagtttcagtgatgcaaatgtgtgc 2940
QY 2652 cctctggttcagctgaaacagtccttggaacttccaanaaccttgaaatagtcctccacagt 2711
|||||
Db 2941 cctctggttcagctgaaacagtccttggaacttccaanaaccttgaaatagtcctccacagt 3000
QY 2712 tgtataaatltygacaattttaggaattttaaactttagatgatcatattgttccattttta 2771
|||||
Db 3001 tgtataaatltygacaattttaggaattttaaactttagatgatcatattgttccattttta 3060
QY 2772 ttcaattttattttgttaatgcaaacagagacttaaatgaactttgatctctgtttta 2831
|||||
Db 3061 ttcaattttattttgttaatgcaaacagagacttaaatgaactttgatctctgtttta 3120
QY 2832 agatattaaaaaacatltgtatctatatcatatagtccttgagagacttagctttcacta 2891
|||||
Db 3121 agatattaaaaaacatltgtatctatatcatatagtccttgagagacttagctttcacta 3180
QY 2892 cactacaggatatgatctccaagttagtcataataaacctgcagagtgattttccagagtg 2951
|||||
Db 3181 cactacaggatatgatctccaagttagtcataataaacctgcagagtgattttccagagtg 3240
QY 2952 ctcgatactgttaattacatctccattagggctgaaagaatgacctacgtttctgtata 3011
|||||
Db 3241 ctcgatactgttaattacatctccattagggctgaaagaatgacctacgtttctgtata 3300
QY 3012 cagctgtgtgtcttltgatgttgtgttactgtacacagaagtggtgacactgaggtctgt 3071

Db 3301 cagctgtgtgtcttltgatgttgttactgttactgtacacagaagtggtgacactgaggtctgt 3360
QY 3072 cgtgtgtgtccgtatgtaaaacctgtgtagccctgcgagtttaagtactgtcttccattcatg 3131
|||||
Db 3361 cgtgtgtgtccgtatgtaaaacctgtgtagccctgcgagtttaagtactgtcttccattcatg 3420
QY 3132 tttaagcttggaatttttctcccatggaatgtgaagtaaaacttaagtgtttgtcatcaat 3191
Db 3421 tttaagcttggaatttttctcccatggaatgtgaagtaaaacttaagtgtttgtcatcaat 3480
QY 3192 aaatggtlaactataaaaaa 3217
Db 3481 aaatggtlaactataaaaaa 3506

RESULT 6
Z93328
ID Z93328 standard; cDNA; 2859 BP.
XX
XX AC Z93328;
XX
DT 04-JUL-2000 (first entry)
XX
DE Mouse staufen cDNA.
XX
KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; mouse; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 325..1788
FT /*tag= a
FT /product= Staufen protein
XX
PN CA2238656-A1.
PN
XX 22-NOV-1999.
PD
XX 22-MAY-1998; 98CA-2238656.
PF
XX 22-MAY-1998; 98CA-2238656.
PR
XX (UYMO-) UNIV MONTREAL.
PA
XX Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI Moulard AJ;
PI
DR WPI: 2000-246924/22.
DR P-PSDB; Y83024.
DR
XX
XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
PT
XX
PS Claim 4; Figure 1c; 96pp; English.
PS
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hSP (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by

CC retroviruses, in particular human immunodeficiency virus.

Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;

Query Match	39.6%;	Score 1273;	DB 21;	Length 2859;
Best Local Similarity	81.2%;	Pred. No. 0;		
Matches 1592; Conservative	0;	Mismatches 275;	Indels 94;	Gaps 6;

OY		1	aactctcgccggagctgcgaggcgcctgaagcgtcttcaagcgtttgcgcgc--ggtctgcgc	58
Db		24	aactctcgccaagctgcgggcccccagcgcctctcaagcgtttgcgtgtcgtcgc	83
OY		59	gtctctcgcgtccgcgtctctlttgaccgcgtcccccccgccggccggccgcgc	118
Db		84	gtctgtgtgcgtcccccctctctcttag-----ccccggcctlgcgcgcgcgc	133
OY		119	tccctcacgccaactccgcctcttcctcccttcgtccctctctctcccttttcc	178
Db		134	ttcgcctccgcacaactccgcctctctccctctctgtgcgcaccttttccctgcgcgtt	193
OY		179	ttcttccctccctccctcctgcgcgcacccgcagacccgcgcgcggggagcagctcoga	238
Db		194	cacttgtcttctcaacctctcctgcgcgcgcaccaagaccgcgcgcgcggagcagcttgy	253
OY		239	gcagcagccagaagtattataaccacttaacctctcagaactgaacaaagaacatgtlt	298
Db		254	ggaa-----	257
OY		299	cctggaagccctcttlttaaaaaagaagcataccctactytagaactaatgcact	358
Db		258	-----gcagccagaagaatatagcttcttacattagctcaatgcact	299
OY		359	gtgcatagaacttggaaaaaaaaccaatgtataagcctgttgaaccttactctcgtatgca	418
Db		300	gtgtgtgaaaccttgaagaanaaaccaatgtataagccctgtgacccctcactctcgtatgca	359
OY		419	gtccacctataactacaacatgtagagaggtgtctatcccccgagttactttaccatt	478
Db		360	gtccacctacagcatatgcatgtcgtltagagtgctctatccccccagataactttaccatt	419
OY		479	ttccagttccaccttacttattatcaagttygaacttctgttggagagacagcaatttaatgy	538
Db		420	ttccagttccaccttacttacttaaccaagtltgaagctctccgttggcgcgacagcagtttaatgy	479
OY		539	caaaggaagaacagaacagcgtgcgaaaaacacgatgtcgtgcgcaaaagcgttgaagatcct	598
Db		480	gaaaggaagaatgtagaaccaaccctgtgaaaaacacgatgtcccttgcgcgtgaagatctt	539
OY		599	gcagaatgagcccccctgcagagagagcttgyaggtgaatgtaaagagaatccgaagaagaaaa	658
Db		540	gcagaatgagccccctgcagaaaggttgyagtltaaatgtaagagaagaagcagaagaagaaaa	599
OY		659	tctcaataaatctgaataaagtcaagtgtttgagattgcaacttaacggaacttgcgt	718
Db		600	cctcaataaatctggaataaagaagccaagtgtttgaaattgcygtlgaagcugaatltgcgt	659
OY		719	gaatttcgaggtgcgcgcgcgcgcgcgcacccacatgtaagaacttgtlgaccaagt	778
Db		660	gaatttttgaggtgcgcgcgcgcgcgcacacacacatgaagaacttgtlgaccaaggt	719
OY		779	ttcgtgttgggagtttgttgggggaaagtgaagggaaaagcaagaagatttccaagaagaaa	838
Db		720	ttcagtttggggaatttgttagggggaaggagaagggaaaagcaagaagatctccaagaagaa	779
OY		839	tgcgcgcacatgctgttctttagagagcttgaagaagttaccgcgcgcctgcgtgcagttgaacy	898
Db		780	tgcgcgcacaggtctgttctttagagcagcttaggaagctgcacacccctccctgtgttgaagcg	839
OY		899	agtaaaagcctagaatcaaaaaagaaaaaacccatagltcaagccacagacaagccacga	958
Db		840	agltgaagcccaagaatcaagaagaagaatlcagccacactgca----agacagccccgga	893

QY	959	ataagggccaaggatcaatccgatctagccgactggcccagatccagcagcaaaaaagga	1018
Db	894	ttatgyccaaggatgaatccttactagtagactgtcacagatccagcagcaaaaaagga	953
QY	1019	gaagagccagagttacacgctctctcacagagcgagcgctcccgcgccgagaggttgt	1078
Db	954	gaagagccagagttacatgctctcttacagaacgaggtcttcacgctgcgagggatttgt	1013
QY	1079	gatcgaggtgaagtttgnaaccacactgcaagaagcagggcaccacaagaaggtggc	1138
Db	1014	gatcgaggtgaagtttggtgcattcacactgcagaagaggtgggtaccataagaagtggc	1073
QY	1139	caagcgcaatgcagcccgagacatgtgtgagatcctctgtttcaagttcccgcgcgca	1198
Db	1074	caagcgtaatgtctgtgagacatgtctgagatcctctgggttccaagttcccgcgcgca	1133
QY	1199	gcccaaccaaacccgcacactcaagtcatgagagagaagacacccataaagaaccagggatgg	1258
Db	1134	gccctgcacaagccagcacactcaaatcagaagagaagaaagactccagtaagaaccaggaagcgg	1193
QY	1259	aagaaaaaglaaccttttttgaaccttggctctggygataaaaaatgygactagtaataaga	1318
Db	1194	aagyaaglaacgttttttgaacctagccctggygataaaaaatgyaactagtaacaaga	1253
QY	1319	ggatgagttcagatgtgcttatcttaagttcatcagcagctgctgtgtgaattcttcccat	1378
Db	1254	cggagagttcagatgtgcttatctttagccatcagcagctgcccagctggaattcttcccat	1313
QY	1379	ggtgcccagaggtcgcccagcgctgtaggagttagtcaaggacatcacaccaagaattttac	1438
Db	1314	ggtgcccaggaagttgcccagcgctgtcggttagtcaaggacacaccacaagaattttcac	1373
QY	1439	caaggcagctccgaaatcctctgccaaagcgcaacggtaaactgcattgtagccggaggttgt	1498
Db	1374	caaggcagctcccaaatcctctgccaaagcgcaacggtaaactgcattgtagccggaggttgt	1433
QY	1499	gtatgggggcacctcgcccacagccgagaccattttaagaataacatcctctcaggcca	1558
Db	1434	gtacgggggcacctcgcccacagccgagaccattttaagaagtaacatcctctcaggcca	1493
QY	1559	cgtaccatgtagacctctcacagagaccctctgagcaactgtagctatcttccagagttcca	1618
Db	1494	cgtaccatgtagacctcgccactagaccctctgagcaactgtagctatcttccagagttcca	1553
QY	1619	gggattccaggttgtaatacaaaagacttccccaaaaacaagaacgaatttgtatctct	1678
Db	1554	gggattccaggttgtaatacaaaagallttcccaagaacaagaacgagttgtatctct	1613
QY	1679	tatcaattgctcctctcagccaccctctgatacagccatggtatcggcaagatgtgagtc	1738
Db	1614	catcaactgtctctcacagccgctctctgtagtcatcagtcgcatcgcaagatgtgagtc	1673
QY	1739	ctgcatgatatagtcgctgcgtgaacatcttaagttgtctgtgagttggaaccaaaag	1798
Db	1674	ctgcatgatatagtcgctgcactgaacattttaagctgctgtgagttggaaccaagag	1733
QY	1799	tacagagatgccaagaacaggaacggaaccaatgtctgtgtgtggyaggtgtctgaacctt	1858
Db	1734	cacagagatgccaagaacaggaacggaatggaaccaatgtctgaagctgtcggyaggtgtctgaacctt	1793
QY	1859	ttctggtccatgaaaccattataaaaatcccaacatatatactgaaaaatactga-aactgctt	1917
Db	1794	ttctggtccacaacacattataaaa-cccaacatatatactgaaaaatactgaaactgtctt	1852
QY	1918	tgaaaatttggaatttctgatacctccagttggyccgagaga	1958
Db	1853	tgaaaatttggaatatctgataactccagttggycccaagaca	1893

RESULT	7
X39991	
ID	X39991 standard; DNA; 1127 BP.
XX	

AC X39991;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
prostate cancer; ss.
XX
OS Homo sapiens.
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 625-626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;

Query Match 34.4%; Score 1107.2; DB 20; Length 1127;
Best Local Similarity 99.4%; Pred. No. 1.3e-272;
Matches 1120; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 520 gagacagcaatttaatgcaagaagacaagcagctgcgaaacacgactgctg 579
Db 1 gagacagcaatttaatgcaagaagacaagcagctgcgaaacacgactgctg 60

QY 580 ccaagcgttgagatcctgcagaaatgagccctgcagagagagctgaggtgaatgaa 639
Db 61 ccaagcgttgagatcctgcagaaatgagccctgcagagagagctgaggtgaatgaa 120

QY 640 gagaatccgaagaagaataatctgaataatctgaataagtcagtgtttgagattgcac 699
Db 121 gagaatccgaagaagaataatctgaataatctgaataagtcagtgtttgagattgcac 180

QY 700 ttaacggaacttgctgtgtaatttcgaggtgcccgggagagtggtgccacccacatga 759
Db 181 ttaacggaacttgctgtgtaatttcgaggtgcccgggagagtggtgccacccacatga 240

QY 760 agaacttgtgaccaaggttcggttggtggagttgtggtggaagtgaagggaagaagca 819
Db 241 agaacttgtgaccaaggttcggttggtggagttgtggtggaagtgaagggaagaagca 300

QY 820 agaagattcaagaanaatgcgcacatagctgttcttgaggagctgaagaagttaccgc 879
Db 301 agaagattcaagaanaatgcgcacatagctgttcttgaggagctgaagaagttaccgc 360

QY 880 cctgtcctgcagttgaaacgagtaaaagcctagaatcaaaaaaagaacaacacatagtica 939
Db 361 cctgtcctgcagttgaaacgagtaaaagcctagaatcaaaaaaagaacaacacatagtica 420

QY 940 agccacagacaagccacagaatatgcccagggatccaatccgattagccgactggcccaga 999
Db 421 agccacagacaagccacagaatatgcccagggatccaatccgattagccgactggcccaga 480

QY 1000 tcacagcagcaaaaaaaggagaaggagccagagttacacgctcctcacagagcaggtctcc 1059
Db 481 tcacagcagcaaaaaaaggagaaggagccagagttacacgctcctcacagagcaggtctcc 540

QY 1060 cgcgcgcgagggagtttgtatgcaggtgaaggttggaaaaaccacactgcagaagaagcgg 1119
Db 541 cgcgcgcgagggagtttgtatgcaggtgaaggttggaaaaaccacactgcagaagaagcgg 600

QY 1120 gcaccacaagaaggtgcccagcgcaatgcagccgagacaatgtctgagatccttggt 1179
Db 601 gcaccacaagaaggtgcccagcgcaatgcagccgagacaatgtctgagatccttggt 660

QY 1180 tcaaatcccgccagcgccgaccccaaaccccgactcaagtcagagagagaagacacca 1239
Db 661 tcaaatcccgccagcgccgaccccaaaccccgactcaagtcagagagagaagacacca 720

QY 1240 taaagaaaccaggggagtgaagaaagtaaacctttttgaacctggtctggtggatgaa 1299
Db 721 taaagaaaccaggggagtgaagaaagtaaacctttttgaacctggtctggtggatgaa 780

QY 1300 atggactagtaataaagagatgagttcagatgccttatctaagtcacgcagctgc 1359
Db 781 atggactagtaataaagagatgagttcagatgccttatctaagtcacgcagctgc 840

QY 1360 ctgctggaattctcccatgtgcccagaggtgcgcccagctgtagagtagtcaagac 1419
Db 841 ctgctggaattctcccatgtgcccagaggtgcgcccagctgtagagtagtcaagac 900

QY 1420 atcacaccaagaattttaccagggcagctccgaatcctgccaagggcaccggtlaactgcca 1479
Db 901 atcacaccaagaattttaccagggcagctccgaatcctgccaagggcaccggtlaactgcca 960

QY 1480 tgatagcccgagagttgtgtatggygggcacctgcgccacag-ccgagaccattttaag 1538
Db 961 tgatagcccgagagttgtgtatggygggcacctgcgccacagcccgagaccattttaag 1020

QY 1539 aataacatctctcagggccacgtaccacatgacaccttcacagagaccctctgagcaactg 1598
Db 1021 aataacatctctcagggccacgtaccacatgacaccttcacagagaccctntgagcaactg 1080

QY 1599 gactatcttccagagctccagggatccaggttgaatacaagaactt 1645
Db 1081 gactatcttncagagctccagggattncaaggttgaatacaagaactt 1127

RESULT 8
X40002
ID X40002 standard; DNA; 946 BP.
XX
AC X40002;
XX
DT 02-JUL-1999 (first entry)
XX

PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Odata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.

XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers

PS Claim 67; Page 629; 787pp; English.

XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

SO Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

Query Match 23.5%; Score 755.6; DB 20; Length 844;
Best Local Similarity 96.3%; Pred. No. 5.3e-183;
Matches 812; Conservative 0; Mismatches 26; Indels 5; Gaps 4;

QY 682 aagtgtttgagattgcacttaaacggaactgctgtgtaatttcgaggtgcccgggaga 741
DB 1 aagtgtttgagattgcacttaaacggaactgctgtgtaatttcgaggtgcccgggaga 60
QY 742 gtggcccaaccacatgaagaacttltgaccaagtttcggttgggagttgtgggg 801
DB 61 gtggcccaaccacatgaagaacttltgaccaagtttcggttgggagttgtgggg 120
QY 802 aagtggaagggaagaagaatttcaagaataatgcgcatagtctgtcttgagg 861
DB 121 aagtggaagggaagaagaatttcaagaataatgcgcatagtctgtcttgagg 180
QY 862 agctgaagaagttaccgcccctgctgcagttagaacgagtaaaagcctagaatcaaaaaga 921
DB 181 agctgaagaagttaccgcccctgctgcagttagaacgagtaaaagcctagaatcaaaaaga 240
QY 922 aaacaaaccatagtcaagccacagacaagaagcccaataatgccaagggatcaatccga 981
DB 241 aaacaaaccatagtcaagccacagacaagaagcccaataatgccaagggatcaatccga 300
QY 982 ttaccgaactggccagatccagcagcgaacaaagaaggaagagccagagtagcacgctcc 1041
DB 301 ttaccgaactggccagatccagcagcgaacaaagaaggaagagccagagtagcacgctcc 360
QY 1042 tcacagagcagagcctccgcgcgcagaggaagtttgtatgcagtgaaagttcgaaacc 1101
DB 361 tcacagagcagagcctccgcgcgcagaggaagtttgtatgcagtgaaagttcgaaacc 420
QY 1102 acactgcagaagaagcggcaccacaagaaggtgtgccaagcgaatgcagcgagaaca 1161
DB 421 acactgcagaagaagcggcaccacaagaaggtgtgccaagcgaatgcagcgagaaca 480

QY 1162 tgctggagatccttggttccaagtcccgccagcgcaagcccaaacccgcactcaagt 1221
DB 481 tgctggagatccttggttccaagtcccgccagcgcaagcccaaacccgcactcaagt 540
QY 1222 cagaggaagaagacacccataaagaacccagggagtgaaagaagaagtaacctttttgaac 1281
DB 541 cagaggaagaagacacccataaagaacccagggagtgaaagaagaagtaacctttttgaac 600
QY 1282 ctgctctgggagtgaataatgagactagtaataaagagatgattcagagtcctatc 1341
DB 601 ctgctctgggagtgaataatgagactagtaataaagagatgattcagagtcctatc 660
QY 1342 taagtcatacgaagctgctgtggaattcttcccatggtgccgaggtgcgccagctg 1401
DB 661 taagtcatacgaagctgctgtggaattcttcccatggtgccgaggtgcgccagctg 720
QY 1402 tagg-agttagtcagaacatcacaccaagaattttaccagggcagctccgaatcctgcc 1460
DB 721 taggaggttagtcagaacatcacaccaagaattttaccagggcagctccgaatcctgcc 779
QY 1461 aagggcaagcgttaactgccaatgatacccgagagttgtgtatggygcacctgcgccaca 1520
DB 780 anggcncng--tactgcattgatagcc--anagttgtgtalttgygcancnttgcgccca 836
QY 1521 gcc 1523
DB 837 ggc 839

RESULT 10
X39992
ID X39992 standard; DNA; 773 BP.
XX
AC X39992;

DT 02-JUL-1999 (first entry)

XX Prostate cancer associated gene.

XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.

XX Homo sapiens.

XX
PN WO9904265-A2.

XX
PD 28-JAN-1999.

XX
PF 15-JUL-1998; 98WO-US14679.

XX
PR 22-JUN-1998; 98US-0102322.

XX
PR 17-JUL-1997; 97US-0896164.

XX
PR 10-OCT-1997; 97US-0061599.

XX
PR 10-OCT-1997; 97US-0061765.

XX
PR 10-OCT-1997; 97US-0948705.

XX
PR 11-OCT-1997; 97GB-0021697.

XX (LUDW-) LUDWIG INST CANCER RES.

XX
PI Chen Y, Gout I, Gure A, O'Hare M, Odata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

XX
DR WPI; 1999-132448/11.

XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers

XX
PS Claim 67; Page 626; 787pp; English.

XX
CC The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

XX
SQ Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

Query Match 23.4%; Score 753.2; DB 20; Length 773;
Best Local Similarity 99.1%; Pred. No. 2.1e-182;
Matches 766; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 874 taccgcccctgctgcaagttgaacagtagaagcctagaatcaaaaaaacaaccaca 933
Db 1 taccgcccctgctgcaagttgaacagtagaagcctagaatcaaaaaaacaaccaca 60
QY 934 tagtcaagccacagacaagccagaatatgcccaggagatcaatccgattagccgactgg 993
Db 61 tagtcaagccacagacaagccagaatatgcccaggagatcaatccgattagccgactgg 120
QY 994 cccagatccagcagcgcaaaaaagagagaagcagagtacacgctcctcacagagcgag 1053
Db 121 cccagatccagcagcgcaaaaaagagagaagcagagtacacgctcctcacagagcgag 180
QY 1054 gcctcccgcgccgcaaggagttgtgatgcaggtgaaggttgaaaccacactgcagaag 1113
Db 181 gcctcccgcgccgcaaggagttgtgatgcaggtgaaggttgaaaccacactgcagaag 240
QY 1114 gaacgggcaccaacaagaagtggtgcacaagcgcaatgcagccgagaacatgtctgagatcc 1173
Db 241 gaacgggcaccaacaagaagtggtgcacaagcgcaatgcagccgagaacatgtctgagatcc 300
QY 1174 ttggtttcaaaagtcctcgagcgagcagccaccaaaaccgcgactcaagtcagagagagaaga 1233
Db 301 ttggtttcaaaagtcctcgagcgagcagccaccaaaaccgcgactcaagtcagagagagaaga 360
QY 1234 cacccataaagaacaagcagggatgtgaagaagaagtaaccttttgaacctggtctggg 1293
Db 361 cacccataaagaacaagcagggatgtgaagaagaagtaaccttttgaacctggtctggg 420
QY 1294 atgaaaatgagctagtaataaagagatgagttcagagatgcctatatcaatcatcagc 1353
Db 421 atgaaaatgagctagtaataaagagatgagttcagagatgcctatatcaatcatcagc 480
QY 1354 agctgcctgctggaatttcccatgtgtcccgaggtgcgcccagctgtagagttagtc 1413
Db 481 agctgcctgctggaatttcccatgtgtcccgaggtgcgcccagctgtagagttagtc 540
QY 1414 aaggacatcacaccaagaattttaccagggcagctccgaatccctgccaaggccaaggtaa 1473
Db 541 aaggacatcacaccaagaattttaccagggcagctccgaatccctgccaaggccaaggtaa 600
QY 1474 ctgcatgatagcccgagaggtgtgtgtatgggggcaacctgcgccacag-ccgagaccatt 1532
Db 601 ctgcatgatagcccgagaggtgtgtgtatgggggcaacctgcgccacagcccgagaccatt 660
QY 1533 ttaagaataacatctcttcagggcagctaccccatgaccttcacagagaccctctgg 1592
Db 661 ttaagaataacatctcttcagggcagctaccccatgaccttcacagagaccctctgg 720
QY 1593 caactgactatcttccagaggtccagagattccaggttgataatacaagactt 1645
Db 721 caactgactatcttccagaggtccagagattccaggttgataatacaagactt 773

RESULT 11
X40001/c
ID X40001 standard; DNA; 825 BP.
XX

AC X40001;

DT 02-JUL-1999 (first entry)

DE Prostate cancer associated gene.

KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.

OS Homo sapiens.

PN WO9904265-A2.

PD 28-JAN-1999.

PE 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfreundschuh M, Sahn U, Scanlan MJ, Stockert E;

PI Tureci O;

XX WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 629; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically

CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the

CC interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in

CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides

CC which are cancer associated antigen precursors expressed in human breast

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

XX
SQ Sequence 825 BP; 289 A; 152 C; 151 G; 228 T; 5 other;

Query Match 17.1%; Score 551; DB 20; Length 825;
Best Local Similarity 95.9%; Pred. No. 8e-131;
Matches 585; Conservative 0; Mismatches 23; Indels 2; Gaps 2;

QY 2600 ttttattcatgactaaccagagagtttcagtgatgcaaatgtgtgcccctgtgt 2659

Db 791 TTTAATTTCCCTGGCTAAACCCAGGNCAGTTCAGTGATGCAAAATGGGGCCNTCTGT 732

QY 2660 tcagctgaaa-cagtcctgagacttcaaaaaccttgataaagctccacagtggtataa 2718

Db 731 TCAGCTGAACACGACTCGACTTTCAAAACCTTGAATAAGTCTCCCCAGTGTATAA 672

QY	2719	attggacaatttaggaattttaactltagatgatcatltaggttccatttcatltaatt	2778
Db	671	ATTGGACCAATTAGGAATTTTAAACCTTAGATGATCATTTGGTCCATTTTATTNTTCATT	612
QY	2779	ttatatttgttaatgcaaacagac-ttaatgaactltagctctgtltaagaatta	2837
Db	611	TTTATTTTTTGAATGCAAAACAGACCTTAAATGAACCTTTGATCTCTGTTTTAAAGATTA	552
QY	2838	ttaaaaaacattgtgtatctatatacatatgctctltaggaccttagcttccactacac	2897
Db	551	TTAAAAAACATGTGTATCTATACATATGGCTCTTGAGGACCTTAGCTTCCACTACACTAC	492
QY	2898	aggatatgatctccatgtltagtccataataaacctgcagagtgatlltccagagtgcgat	2957
Db	491	AGGATATGATCTCCATGTAGTCCATATTAACCTGCAGAGTGAATTTTCCAGAGTGCCTCAT	432
QY	2958	actgttaattacatctccatlagggctgaaaagaatgaactacgtllctglatacagctg	3017
Db	431	ACTGTTAATTACATCTCCATTAGGGCTGAAAGAATGACCTACGTTTCTGTATACAGCTG	372
QY	3018	tgltgctttagatgtlgtlgttactgtacacagaagltgtgcactgaaggctctgcgtg	3077
Db	371	TGTTGCTTTTGAATGTTGTGTACTGTACACAGAAGTGTGTGCACTGAAGCTCTGCGTGTG	312
QY	3078	gtccglatgaaaaacctggtagccctgcgaggtlaagtlactgtcttccatctcatlgtttag	3137
Db	311	GTCCGATGGAAGCCCTGTAGTACCCCTGCCAGTTAAGTACTGCTTCCATTCAATTGTTACG	252
QY	3138	ctggaattttcttcccccatggaatgtgaagtaaaacttaagtgttgtlcatlcaataaatg	3197
Db	251	CTGGAATTTTCTCCCATGGAATGTAGTAATAAACTTAAGTGTTTGTCAATCAATAAATGG	192
QY	3198	taataactaaa 3207	
Db	191	TAATACTAAA 182	
RESULT	12		
QY	X39993/c		
ID	X39993	standard; DNA; 794 BP.	
AC	X39993;		
DF	02-JUL-1999	(first entry)	
XX		Prostate cancer associated gene.	
KW		Cancer associated antigen; diagnosis; research; treatment; human;	
KW		breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KW		prostate cancer; ss.	
XX	OS	Homo sapiens.	
XX	PN	WO9904265-A2.	
PD	28-JAN-1999.		
XX			
PF	15-JUL-1998;	98WO-US14679.	
XX			
PR	22-JUN-1998;	98US-0102322.	
PR	17-JUL-1997;	97US-0896164.	
PR	10-OCT-1997;	97US-0061599.	
PR	10-OCT-1997;	97US-0061765.	
PR	10-OCT-1997;	97US-0948705.	
XX	11-OCT-1997;	97GB-0021697.	
PA	(LUDW-)	LUDWIG INST CANCER RES.	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;		
PI	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;		
PI	Tureci O;		
XX			

DR	WPI; 1999-132448/11.
XX	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Claim 67; Page 626; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
SQ	Sequence 794 BP; 273 A; 141 C; 152 G; 206 T; 22 other;
Query Match	15.6%; Score 501.4; DB 20; Length 794;
Best Local Similarity	96.3%; Pred. No. 3.5e-118;
Matches 517; Conservative	0; Mismatches 18; Indels 2; Gaps 1;
QY	2673 tcctggaacttcaaaaacctgtgaataagtcctccc--acagtgtataaatggacaattt 2730
Db	718 TCCTGGACTTTCAAAACCCTTGAAATAGTCTCCCCNCCAGTTGNATAAATTGGCCMATTT 659
QY	2731 aggaattttaaaacttagatgatcatcatttgltccattttlatltaattttatttgtt 2790
Db	658 AGGAATTTTAAACNTTAGATGTTCAATTGGTTCATTTTTATTTCAATTTTATTTTGGT 599
QY	2791 aaatgcaaacaggacttaaatgatgaacttgatctctgttttaaagaattataaaaaacatty 2850
Db	598 AATGCAACAACAGNCTTAATGAACCTTGATCTCTGTTTTAAAGAATTATTA AAAAACATTG 539
QY	2851 tgcatacatatatgtgctcttgagagacttagctttcactactacacagatatgatctc 2910
Db	538 TGNATCTATACATATGGCTCTTGAGAGACTTAGCTTTCACCTACACTACAGATATGATCTC 479
QY	2911 catgtagtccaataaaacctgcagagtgaatttccagagtgctcgatactgtttaattaca 2970
Db	478 CATGTAGTCCATATATAACCTGCAGAGTGATTTTCCANAGTGCTCGATACGTTAATTACA 419
QY	2971 tctccatagggctgaaaagaatgacctacglltctgtatacagctgtgtcclttgat 3030
Db	418 TCTCCMTTAGGGCTGAAAAAGAAATGACCTACGTTCTGTATACANCCTGTGTGCTTTTGAT 359
QY	3031 gttgtgtactgtacacagaagtgtgtgcaactgaagctctgcgtgtgtgccgtatgaa 3090
Db	358 GTTGTGNTACTGNACACAGAAAGTGTGTCNCCTGAGGCTCTGCGTGTGCCGTATGAAA 299
QY	3091 acctgtagccctgcgagtltaagtactgtctccattcatgtltaagctcggaaatttctc 3150
Db	298 GCCTGGTAGCCCTCGAGAGTTAAGTACTGCTTCCATTTCATTTGTTACGCGGAATTTTCT 239
QY	3151 ccccatggaatgttaagtaaaacttaagtgtttgtcatcataaatgtgtaataactaaa 3207
Db	238 CCCCATGCAATGTAGTAAACCTTAAGTGTGTTCATCATCATNAATGGTATACTAAA 182
RESULT	13
ID	Z79981/c
AC	Z79981 standard; cDNA; 660 BP.
XX	Z79981;
XX	

DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
XX hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801;
XX
PA (FARB) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
XX Claim 1; Page 158; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
XX Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other;
XX
SO
Query Match 13.5%; Score 433.8; DB 21; Length 660;
Best Local Similarity 92.4%; Pred. No. 5.5e-101;
Matches 495; Conservative 0; Mismatches 36; Indels 5; Gaps 4;
QY 1026 ccagagtacagctcctccacagagcagcctcccgccgcgcagggagttgtatgcag 1085
DB 531 CCAAGGTACACGTTCTTCANGAAGCNAGCCTTCCGGNGCGNAGGAAGTGTGTATNCAG 472
QY 1086 gtgaaggttggaaacacactgcagaagcagcggcaccacaagaaggtggccaagcgc 1145
DB 471 GTNAAAGGTTGAACCCCATTTGCANAAAGGA--CGGGCACCAAGAGGTGGCCAAG-GC 415
QY 1146 aatgcagccgagaacatgctgagatccttggttccaagtcccgagcgcgcagccacc 1205
DB 414 AATGCAG-CGAGANCAATGTTGAAGATCCTGGGTTTCAAGTCCCGCAGCGCCAG-CCACC 357
QY 1206 aaaccgcactcaagtcagagagagaagacaccataaagaacaggggatggaanaaa 1265
DB 356 AAACCGGCACTCAAGTCAGAGAGAGAACCCCATATAAGAAACCAAGGGATGGAAGAAA 297
QY 1266 gtaacctttttgaacctgctctggtgagtaaaatggtgactagtataaagagatgag 1325
DB 296 GTAACTTTTGTGAACCTGGCTCTGGGGATGAATAATGGGACTAGTAATAAGAGGATGAG 237
QY 1326 ttcagatgccttatctaatgcatcagcagctgcctgctggaattcttcccatggtgcc 1385
|||||

DB 236 TTCAGGATGCCCTTATCTAAGTCAATCAGCAGCTGCCCTGCTGGAATTCCTCCATGTGCC 177
QY 1386 gagctgcgccagcgtctgagagtagtcaagagacatcacaccaaagatttaccaggca 1445
DB 176 GAGCTGCGCCAGCGCTGTAGAGAGTTAGTCAAGGACATCACACCAAGAATTTTACAGGGA 117
QY 1446 gctccgaatcctctgccaagggccaagcgtactgcatgatagcccgagaggtgtgtatgg 1505
DB 116 GCTCCGAATCCTGCCAAGGCCACGGTACTGCGCATGATAGCCCGAGAGTGTGTATGGG 57
QY 1506 ggcacctgcgccacagcgcagaccattttaaagaataacatctcttcaggccacgt 1561
DB 56 GGCACCTCGCCACAGCCGAGACCATTTTAAAGAATAACATCTCTTACAGCCACGT 1
RESULT 14
280269
ID 280269 standard; cDNA; 727 BP.
XX
AC 280269;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
XX hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
XX Claim 15; Page 273; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
XX Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;
XX
SO
Query Match 12.8%; Score 412.2; DB 21; Length 727;
Best Local Similarity 95.5%; Pred. No. 1.9e-95;
Matches 444; Conservative 0; Mismatches 17; Indels 4; Gaps 2;
QY 463 ggtacttttaccattccagttccacccttacttatacaagtgaacttctgtggag 522
|||||

|||||
Db 1 ggtactttaccatttcacgttcacatttacttalcacagtgaactttctgtggag 60
QY 523 gacagcaatttaatgcaagaagaacagacagctgcgaacacgactgtgtgcc 582
Db 61 gacagcaatttaatgcaagaagaacagacagctgcgaacacgactgtgtgcc 120
QY 583 aagcgttgaggtatcctgcagatgagccctgcagagagctggaagtgaatgaag 642
Db 121 aagcgttgaggtatcctgcagatgagccctgcagagagctggaagtgaatgaag 180
QY 643 aatccgaagaagaataatctcaataaatctgaataagtgtgttgagattgcactta 702
Db 181 aatccgaagaagaataatctcaataaatctgaataagtgtgttgagattgcactta 240
QY 703 aacggaactgcctgtgaatttcgaggtgccccggagagtgccccacacatgaaga 762
Db 241 aacggaactgcctgtgaatttcgaggtgccccggagagtgccccacacatgaaga 300
QY 763 acttgtgacccaaggttccgtgtgggagttgtgggggaaggtgaaggaaagcaaga 822
Db 301 acttgtgacccaaggttccgtgtgggagttgtgggggaaggtgaaggaaagcaaga 360
QY 823 agatttcaaaagaaaatg-ccgcacatagctgttcttgagagctgaagaagttaccgcc 881
Db 361 agatttcaaaagaaaatgccccgcacatagctgttcttgagagctgaagaagttaccgcc 420
QY 882 ctgcctgcagttgaac--gaataagcctagaatcaaaaagaaa 923
Db 421 ctgcctgcagttgaac--gaataagcctagaatcaaaaagaaa 465

RESULT 15
280703/c
ID 280703 standard; cDNA; 708 BP.

XX 280703;
XX 07-APR-2000 (first entry)
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:787.
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:787.
XX
XX Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-1B01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Carino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
PS Claim 15; Page 444; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The

CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 708 BP; 226 A; 127 C; 143 G; 191 T; 21 other;

Query Match 10.2%; Score 327.8; DB 21; Length 708;
Best Local Similarity 82.8%; Pred. No. 6.2e-74;
Matches 515; Conservative 0; Mismatches 83; Indels 24; Gaps 12;

QY 2446 ttctcttattgatataaac--tatataaatttcatattgagaattctc--aat 2500
Db 622 TTTCCNTTNGATGAAATGGCCCTTAATAAAAAATTCCNTTGAATAATTTCCCAATG 563
QY 2501 tgtatctagttaaatagcacagtttgaaaact--gtctgagactgacttatc--aat 2555
Db 562 GGATCCAGTTAAATAGCCCGAGTTGNAAACTTGGTTAAAGCCTGCTTAAACCAATA 503
QY 2556 aatctaaccgac--aaagatcatatccatgtgtatgtgttagacatttattcatgt 2613
Db 502 ATTCNAACCGGCCAAAGTTCAAAACCCNAGGGGAAGNGNTAAGCNATTTTAATTCATG 443
QY 2614 actaacc--aggacagtttcaatgtagcaaatgtgtgcccctc--ggttcaagctgaaa- 2669
Db 442 NCTTACCCAGGCGCCAGTTCCAGTAAGCAAAATTTGGGGGCTTCTGGGTTCAAGCTGAAC 383
QY 2670 cagtcctgactttcaaaaacacttgaa-taagtcctcca-cagttgtataaatgtgacaa 2727
Db 382 CAGTCTGACTTTCCAAAACCTGAATTAAGTCTCCACACAGTTGGATAAATTGNCNA 323
QY 2728 tttagaattttaaactttagatgaca----tttggttcaatttlatlcatlta 2782
Db 322 TTTAGGAATTTTAAACCTTTTAGATGATCCATTGGTCCCATTTTATTTCAATTTTA 263
QY 2783 ttltgttaatgcaaacagacttaaatgaactttagatctctgttttaagattatnaa 2842
Db 262 TTTTGTAAATGCAAAACAGACTTAATGAACCTTTGATCTCTGTTTAAAGATTATTA 203
QY 2843 aaacattgtatctatcatatcatatgctcttgaggactttagcttcaactacacagat 2902
Db 202 AAACATGTGTATCTATACATATGGCTTTGAGGACTTAGCTTCACTACACTACAGAT 143
QY 2903 atgattccatgtagtccatataaaacctgcagagtgaatttccaagagtgtcgatactgt 2962
Db 142 ATGATCTCCATGTAGTCCATATAAACCCTGCAGAGTGATTTTCCAGAGTGCTCGATACTGT 83
QY 2963 taattacatctcatatgaggtgaagaatgaactagcttctgtatatacag-ctgtgtt 3021
Db 82 TAATTACATCTCATTTAGGGCTGAAGAAATGACCCAGTTCTGTATACAGCCTGTGTT 23
QY 3022 gctttgatgtgtgttactgt 3043
Db 22 GCTTTTGATGTGTGTACTGT 1

Search completed: April 5, 2001, 08:30:05
Job time: 59249 sec

alignment_block:
US-09-316-048-11 x 293328 ..

Align seg 1/1 to: 293328 from: 1 to: 2859

```
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
325 ATGTATTAAGCCCGTGAAGCCTCCTCGGATGCAGTCCACCTACAGCTA 374

17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
375 TGGCATGCGTGGAGGTGCTTATCCCCCAGATACTTTTACCCATTTCAG 424

34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
|||||
425 TCCCACCTTACTCTAACAAGTTGAGCTTCCGTTGGCGGACAGCAGTTT 474

51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
475 AATGGAAAGGAAGATGAGACACCCCGTGAACACGATGCCCTGCCCG 524

67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
|||||
525 TGGCGTGAAGACTCTGCAGAGTGAACCCCTGCCAGAAAGTTGGAGGTA 574

84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
|||||
575 ATGGAAGAGAGAGCAGAGAGAAACCTCAATAATCGGAATATAGCCAA 624

101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
625 GTGTTGAAATTGCCGCTGAAGCGGAATTGCCCTGTGAATTTGAGGTGCC 674

117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
|||||
675 CCGGAGAGTGGCCACACCATGAAGAACTTTGTGACCAGGCTTCAG 724

134 aGlyGluPheValGlyGluGlyGlyLysSerLysLysIleSerLys 150
|||||
725 TTGGGAATTTGTAGGGAAGAGAGAAAGCAAGAGATCTCCAAG 774

151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProle 167
|||||
775 AAGAATGCGGCAGGCGCTTCTTGAGCAGCTTAGGAGGCTGCCACCCCT 824

167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnProT 184
|||||
825 CCTGCTGTGAGCGAGCTGAAGCCAGATCAAGAAGAAAGTCAAGCCCA 874

184 hrcysLysThrAlaProAspTyrGlyGlnGlyMetAsnProIleSerArg 200
|||||
875 CTTGCAAGACAGCCCCGGATTATGGCCAGGGGATGAATCCTATTAGTAGA 924

201 LeuAlaGlnIleGlnAlaLysLysGluLysGluProGluTyrMetIle 217
|||||
925 CTTGCACAGATCCAGCAGGCAAAAAAGAGAGAGGCCAGAGTACATGCT 974

217 uLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGlnValL 234
|||||
975 CTTTACAGAAAGAGCTCTTCCACGTCGCGAGGAGTTTGTGATGCAGGTA 1024

234 ysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLysValAla 250
|||||
1025 AGGTTGGGCATCACACTGCAGAAGAGAGTGGGTACCAATAAGAGAGTGCC 1074

251 LysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysValPr 267
|||||
1075 AAGCGTAATGCTGCTGAGAACATATGTCGAGATCCTGGGGTTCAAGTTCC 1124

267 oGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysThrProV 284
|||||
1125 CCAGGCGCAGCCTGCGCAAGCCAGCACTCAAAATCAGAAAGAGAGACTCCAG 1174
```

```
284 alLysLysProGlyAspGlyArgLysValThrPhePheGluProSerPro 300
|||||
1175 TAAAGAAACCGAGAGCGAAGGAAGTAACGTTTTTTGAACCTAGCCCT 1224

301 GLyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetProTy 317
|||||
1225 GGGGATGAAATGGAAGTACTAGTAACAAGAGCAGAGGATTCAAGATGCCCTTA 1274

317 rLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluV 334
|||||
1275 TCTTAGCCATCAGCAGCTGCCAGCTGGAATTCTCCCATGTGTCGGGAAG 1324

334 aLAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThr 350
|||||
1325 TTGCCAGGCTGTCGGGGTTAGTCAAGGACACACCAAGATTTCCACC 1374

351 ArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaAr 367
|||||
1375 AGGCGAGCTCCAATCTCTGCCAAGGCAACGGTAACGTGCATGATAGCCCG 1424

367 gGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLys 384
|||||
1425 AGAGTTGTTGACGGGGGCACCTCGCCACAGCCGAGACCATTTTAAAGA 1474

384 eRAsnIleSerSerGlyHisValProHisGlyProArgThrArgProSer 400
|||||
1475 GTAACATCTCTTCAAGCCACGTAACCCCATGGAACCTCGCACACTAGACCTCT 1524

401 GluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGluTyrLy 417
|||||
1525 GAGCAACTGTACTACCTTTCAGAGAGCCAGGAGATTCCAGGTTGAATACAA 1574

417 sAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleAsnCys 434
|||||
1575 AGATTTTCCCAAGACAACAAGACGAGTGTATCTCTCATCACTGCT 1624

434 eRserGlnProProLeuValSerHisGlyIleGlyLysAspValGluSer 450
|||||
1625 CTTCAACAGCCGCTCTCGTCAGTCATGGCATCGGCAAGAGATGTGAGTCC 1674

451 CysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLeuAs 467
|||||
1675 TGTATGATATGCTGCTGCACCTGAACATTTTAAAGCTGCTGTCGAGTTGGA 1724

467 pGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValSerAlaC 484
|||||
1725 CCAACAGAGCAGCAGATGCCAAGAAGCAGAGAAATGGACCAGTTTCAGCGT 1774

484 ysGlyArgCys 487
|||||
1775 GCGGAGAGGTGC 1785
```

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X90786

seq_documentation_block:

ID X90786 standard; DNA; 3190 BP.

XX AC X90786;

DT 13-JAN-2000 (first entry)

XX DNA encoding human stauflen protein.

DE Human stauflen protein; hstau; protein lysate; lung; kidney; testis;

KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;

KW telomerase assembly; maturation; transport; regulation; interaction;

XX inhibitory agent; antibody; immunoassay; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 345..1835

```
FT      /*tag= a
FT      /product= "Human staufen protein"
FT      /note= "Binds to telomerase RNA"
FT      3172..3177
FT      polyA_signal
FT      /*tag= b
XX      WO9951255-A1.
XX      14-OCT-1999.
XX      06-APR-1999; 99WO-US07533.
XX      06-APR-1998; 98US-0080783.
XX      (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX      Greider CW, Le S;
XX      WPI; 1999-620168/53.
DR      P-PSDB; Y26335.
XX      Human staufen polypeptide useful in methods for identifying telomerase
PT      inhibitors
XX      Claim 13; Page 41-42; 50pp; English.
XX      The present sequence is an isolated DNA encoding human staufen (hstau)
CC      protein. The hstau protein was detected in protein lysates from lung,
CC      kidney, testis and ovary, but not in brain and heart. It binds
CC      specifically to human telomerase RNA and can complex with the
CC      telomerase catalytic subunit. The hstau protein may play a role in
CC      telomerase assembly, maturation, transport and regulation. Interaction
CC      between hstau and telomerase can be inhibited by administering inhibitory
CC      agents to the cell. These telomerase inhibitors can be identified using
CC      hstau protein. Antibodies generated against hstau can be used in various
CC      immunassays.
XX      SQ      Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;
```

```
alignment_scores:
      Quality: 2333.00      Length: 489
      Ratio: 4.922      Gaps: 1
Percent Similarity: 96.933      Percent Identity: 91.002
```

alignment_block:
US-09-316-048-11 x X90786 ..

Align seg 1/1 to: X90786 from: 1 to: 3190

```
1 MetTyrIysProValaspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
366 ATGTATAAGCCTGTGACCCCTTACTCTCGGATGCGGTCCACTATACTA 415
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|:::|||||
416 CAACATGAGAGAGGTGCTTATCCCCGAGGTACTTTACCACTTCCAG 465
34 alProProleuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
|||||
466 TTCACACTTTACTTTATCAAGTGAAGTCTTCTGTGGAGAGACAGCAAT 515
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
|||||
516 AATGGCAAGGAAAGACACAGACAGCGCTGCGAAACACGATGCTGCTCAA 565
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
:|||||
566 AGCGTTGAGATCCTGCAGATGAGCCCTGCGAGAGAGCTGGAAGTGA 615
84 sngIArgGluAlaGluGluAsnLeuAsnLysSerGluIleSerGln 100
|||||
616 ATGGAAGAGAAATCCGAAGAAATCTCAATAATCTGAATAAGTCAA 665
```

```
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
666 GTGTTGAGATTGCACCTTAACCGAAGTTCCTGTGAATTCGAGGTGGC 715
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValServ 134
|||||
716 CCGGAGAGATGGCCCCACCCACATGAGAACTTGTGACCAGCGTTTCGG 765
134 alGlyGluPheValGlyGluGlyLysSerLysLysIleSerLys 150
|||||
766 TTGGGAGATTGTGGGGGAAGGTGAAGGGGAAAGCAAGAAGATTTCAAAG 815
151 LysAsnAlaAlaArgAlaValLeuGlnLeuArgArgLeuProProLe 167
|||||
816 AAAAATGCCGCCATAGCTGTCTTGAGAGAGCTGAAGAAGTTACCGCCCT 865
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnProT 184
|||||
866 GCCTGACAGTTGAACGAGTAAGCCTAGAAATCAAAAAGAAACAAACCCA 915
184 hrcLysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||
916 TAGTCAAGCCACAGACAGACCCAGAAATATGCGCAGGGAGATCAATCCG 965
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTy 215
|||||
966 AGCCGACTGGCCCGATCCAGCAGGCAAAAAAGAGAGAGAGCCAGAGTA 1015
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|
1016 CACGCTCCTCACAGAGCGAGGCTCCCGCGCCGAGGAGATTGTGATGC 1065
232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
1066 AGGTGAAGGTTGGAACACACACTGCAGAAAGAACGGGCACCAACAAGAAG 1115
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 265
|||||
1116 GTGGCCAAAGCGCAATGCAGCCGAGACATGCTGGAGATCCTTGGTTTCAA 1165
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
|||||
1166 AGTCCCGCAGGCGCAGCCACCAACCCGCACTCAAGTCCAGAGAGAAGA 1215
282 hrProValLysLysProGlyAspGlyArgLysValThrPheGluPro 298
|||||
1216 CACCATATAAGAAACACAGGGATGGAAGAAAGTAACCTTTTGAACCT 1265
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
:::
1266 GGCTCTGGGAGTAATAATGGAGTACTAGTAATAAGAGATGAGTTCAGAGAT 1315
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
|||||
1316 GCCTTATCTAAGTCATCAGCAGCTGCCCTGCTGGAATTTCTCCCATGTGC 1365
332 roGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisHisThrLysAsp 348
|||||
1366 CCGAGGTGCGCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGAT 1415
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIle 365
|||||
1416 TTTACCAGGGCAGCTCCGAATCTCGCAAGGCCACGGTAACCTGCATGAT 1465
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382
|||||
1466 AGCCCGAGAGTGTGTATGGGGCACCCTGCCACACAGCCGAGACCAATT 1515
382 eLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
|||||
1516 TAAAGATATACATCTCTTCAGGCCACGTACCCCATGGAACCTTCACGAGA 1565
```


399 ProSerGIuInLeuTYrTYrLeuSerArgAlaGlnGlyPheGlnValG1 415
|||||
1566 CCCTCTGAGCAACTGACTATCTTCCAGAGTCCAGGATTCAGGTTGA 1615
415 uTYrLYsAsPheProLYsAsnAsnLYsAsnGluCysValSerLeuIleA 432
|||||
1616 ATACAAGACTTCCCAAAACACACAGACGAATTGTATCTCTTATCA 1665
432 sNCysSerSerGIuInProProLeuValSerHisGlyIleGlyLYsAspVal 448
|||||
1666 ATTGCTCTCTCAGCCACTCTGATCAGCCATGTATCGGCAAGATGTG 1715
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLYsLeuLeuSerGI 465
|||||
1716 GAGTCTGCGCATGATATGCGCTGCGCTGAACATCTTAAGTTGCTGTGA 1765
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
|||||
1766 GTTGACCAACAAGTACAGAGATGCCAAGACAGAAACGACCAATGT 1815
482 eAlaCysGlyArgCys 487
|||
1816 CTGTGTGTGGAGGTGC 1832

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:293324

seq_documentation_block:
ID 293324 standard; cDNA; 3142 BP.

XX 293324;

XX 04-JUL-2000 (first entry)

XX Human staufen cDNA.

XX Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX Homo sapiens.

XX Key Location/Qualifiers
FT CDS 288..1193
FT CDS /*tag= a
FT CDS /product= Staufen protein

XX CA2238656-A1.

XX 22-NOV-1999.

XX 22-MAY-1998; 98CA-2238656.

XX 22-MAY-1998; 98CA-2238656.

XX (UYMO-) UNIV MONTREAL.

XX Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;
PI Moulard AJ;

XX WPI: 2000-246924/22.

XX P-PSDB; Y83108.

XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)

XX Claim 4; Figure 1b; 96pp; English.

XX Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to

CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.

XX Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

alignment_scores:
Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x 293324 ..

Align seg 1/1 to: 293324 from: 1 to: 3142

1 MetTYrLYsProValAspProHisSerArgMetGlnSerThrTYrSerTY 17

309 ATGTATAAGCCGTGTGACCCCTTACTCTCGATGCAGTCCACCTATACTA 358

17 rGlyMetArgGlyGlyAlaTYrProProArgTYrPheTYrProPhePro 34

359 CAACATGAGAGGAGGTGCTTATCCCCGAGGTACTTTTACCCTATTCAG 408

34 alProProLeuLeuTYrGlnValGluLeuSerValGlyGlnGlnPhe 50

409 TTCCACCTTACTTATCAAGTGAACCTTCTGTGGAGAGACAGCAATT 458

51 AsnGlyLYsGlyLYsMetArgProProValLYsHisAspAlaProAla 67

459 AATGCCAAAGAAAGACAAGACAGCGCTGCAGAAACAGATGCTGCTGCAA 508

67 gaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84

509 AGCGTTGAGGATCCTGCAGATGAGCCCTGCCAGAGAGGCTGAGGTGA 558

84 snGlyArgGluAlaGluGluGluAsnLeuAsnLYsSerGluIleSerGln 100

559 ATGCAAGAGATCCGACAGAAAGAAATCTCAATAAATCTGAATAAGTCAA 608

101 ValPheGluIleAlaLeuLYsArgAsnLeuProValAsnPheGluValAl 117

609 GTGTTGAGATTGCACCTTAACGGAACCTTGCCGTGTAATTCGAGGTGCC 658

117 aArgGluSerGlyProProHisMetLYsAsnPheValThrArgValSer 134

659 CCGGAGAGTGGCCACCCACATGAAGAACTTTGTGACCAAGGTTTCGG 708

134 alGlyGluPheValGlyGluGlyGlyLYsSerLYsLYsIleSerLYs 150

709 TTGGGAGGTTTGTGGGGGAAGGTGAAGGGAAGCAAGAGATTTCAAAG 758

151 LYsAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProle 167

759 AAAAATGCCCGCATAGCTGTCTTGAGGAGCTGAAGAAGTTACCGCCCT 808

167 uProAlaValGluArgValLYsProArgIleLYsLYsLYsSerGlnProT 184

809 GCCTGCAGTTGACGAGTAAAGCCTAGAAATCAAAAAAGAAACAAACCCA 858

184 hrcLYsLYs.....ThrAlaProAspTYrGlyGlnGlyMetAsnProIle 198

859 TAGTCAAGCCACAGACAAAGCCAGAAATATGCGCAGGGATCAATCGATT 908

199 SerArgLeuAlaGlnIleGlnGlnAlaLysGluLysGluProGluTyr 215
|||||
909 AGCCGACTGCCCATGATCCAGCAGCAAAAAAGAGAGAGAGCCAGAGTA 958
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
| |||||
959 CACGCTCCTCACAGACGAGGCTCCCGCGCCGACGAGGAGTTGTGATGC 1008
232 lnValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
1009 AGGTGAAGGTTGGAAACACACACTGCAGAGAGAACGGGCACCAACAAGAAG 1058
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 265
|||||
1059 GTGGCCAAGCCCAATGCACCGCAGAACATGCTGGAGATCCTGGTTCAA 1108
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
|||||
1109 AGTCCCGCAGCGCAGCCACCAACCCGCACTCAAGTCAGAGAGAGAAGA 1158
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|||||
1159 CACCCATTAAGAAACACAGGGGATGGAAGAAAGTAACCTTTTGAACCT 1208
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
::: |||||
1209 GGCTCTGGGGATGAATAATGGGACTAGTAATAAGAGAGATGAGTTCAGGAT 1258
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
|||||
1259 GCCTTATCTAAGTCATCAGCAGCTGCTGCTGGAATTCTTCCCATGCTGC 1308
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
|||||
1309 CCGAGGTCGCCAGGCTGTAGAGTAGTCAAGGACATCACACCAAGAT 1358
349 pheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIl 365
|||||
1359 TTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGGTAACCTGCCATGAT 1408
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382
|||||
1409 AGCCCGAGAGTTGTTGATGGGGCACCTCGCCACAGCCGAGACCATTT 1458
382 euLysSerAsnLysSerGlyHisValProHisGlyProArgThrArg 398
|||||
1459 TAAAGATTAACATCTCTTCAGGCCACGTAACCCATGACCTCTCAGAGAGA 1508
399 proSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGl 415
|||||
1509 CCCTCTGAGCAACTGGACTATCTTCCAGAGTCCAGGATTCAGGTTGA 1558
415 uTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleA 432
|||||
1559 ATACAAGAGACTCCCAAAAACACACAGACGAATTGTATCTCTTATCA 1608
432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
|||||
1609 ATTGCTCTCTCAGCCACCTCTGATCAGCCATGCTATCGGCAAGGATGTG 1658
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGl 465
|||||
1659 GAGTCTGTCATGATATGCTGCGCTGAACATCTTAAAGTTGCTGTGA 1708
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
|||||
1709 GTTGACCAACAAGTACAGAGATGCCAAGAACAGAAACGAGACCAATGT 1758
482 erAlaCysGlyArgCys 487
|:::|
1759 CTGTGTGTGGAGGTGC 1775

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z93323
seq_documentation_block:
ID Z93323 standard; cDNA; 3217 BP.
XX
AC Z93323;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human staufen cDNA.
XX
KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 363..1853
FT FT /*tag= a
FT FT /product= Staufen
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L;
PI Moulard AJ;
XX
DR WPI: 2000-246924/22.
DR P-PSDB; Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC Z93322.
XX
SQ Sequence 3217 BP; 889 A; 796 C; 742 G; 790 T; 0 other;

alignment_scores:
Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
us-09-316-048-11 x Z93323 ..

PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83023.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
XX
PS Claim 4; Figure 1a; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

alignment_scores: Quality: 2332.00 Length: 489
 Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x Z93322 ..

Align seg 1/1 to: Z93322 from: 1 to: 3506

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
673 ATGTATTAAGCCTGTGACCCCTTACTCTCGATGCAGTCCACCTATACTA 722
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|::|
723 CAACATGAGAGAGAGGTGCTTATCCCCGAGGTACTTTAACCATTTCAG 772
34 alProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
|
773 TTCACCTTTACTTATCAAGTGAACCTTCTGTGGAGACACACAATT 822
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
|
823 AATGGCAAGGAAAGACACAGCGCTGCGAAACACGATGCTGTGCCAA 872
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
|
873 AGCGTTGAGGATCCTGCAGAAATGAGCCCTGCCAGAGAGCTGAGGTGA 922
84 sngLYArGGLuAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
|
923 ATGGAAGAGAAATCCGAAGAAATACTCAATAATCTGAATAAGTCAA 972

101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|
973 GTGTTTGAGATTCACACTTAACCGAAGCTTGCTGTGAATTTCCAGGTGGC 1022
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValServ 134
|
1023 CCGGAGAGTGGCCACCCACACATGAAGAACTTGTGACCAAGGTTTCGG 1072
134 aGlyGluPheValGlyGluGlyLysSerLysIleSerLys 150
|
1073 TTGGGAGTGTGGGGAAGGTGAAGGAAAGCAAGAGATTCCAAG 1122
151 LysAsnAlaAlaArgAlaValLeuGlnLeuArgArgLeuProProle 167
|
1123 AAAAATGCCGCCATAGCTGTCTTGAGAGAGCTGAAGAGATTACCGCCCT 1172
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnProt 184
|
1173 GCCTGCAGTTGAACGAGTAAGCCTTAGAATCAAAAAGAAACAAACCCA 1222
184 hrCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProile 198
|
1223 TAGTCAAGCACAGACAGAGCGCTCCCGCGCCGAGGAGTTGTGATGC 1272
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTy 215
|
1273 AGCCGACTGGCCAGATCCAGACAGCAAAAAAGAGAAGAGCCACAGATA 1322
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|
1323 CACGCTCCACACAGAGCGAGCGCTCCCGCGCCGAGGAGTTGTGATGC 1372
232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
|
1373 AGGTGAAGTTGGAAACACACACTGCAGAGAAGAACGGGCCACCAACAAGA 1422
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPhe 265
|
1423 GTGGCCAGCGCAATGCAGCCGAGACATGCTGAGATCCTTGTTTCAA 1472
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
|
1473 AGTCCCGCAGCGCAGCCACCAACCCGCACTCAAGTCAAGAGAGAAGA 1522
282 hrProValLysLysProGlyAspGlyArgLysValThrPheGluPro 298
|
1523 CACCACATAAGAAACACAGGGGATGGAAGAAAGTAACCTTTTGAACCT 1572
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgme 315
|
1573 GGCTCTGGGGATGAAATGGGACTAGTAATAAGAGGATGAGTTCCAGGAT 1622
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
|
1623 GCCTTATCTTAAGTCATCAGCAGCTGCTGCTGAATTCTTCCCATGTGC 1672
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
|
1673 CCGAGTGGCCAGGCTGTAGAGATTAGTCAAGGACATCACACCAAGAT 1722
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIl 365
|
1723 TTACACAGGGCAGCTCCGAATCTGCCAAGGCCACGTAACCTGCCATGAT 1772
365 eAlaArgGluLeuLeuTyrGlyThrSerProThrAlaGluThrIleL 382
|
1773 AGCCCGAGAGTTGTGTATGGGGCACCTCGCCACAGCCGAGACCATTT 1822
382 euLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
|
1823 TAAAGATTAATCTCTTCAGGCCACGTAACCCATGGACCTCTCACGAGA 1872
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGl 415

```
|||||
1873 CCCTCTGAGCAACTGGACTATCTTCCAGAGTCCAGGATTCAGGTGA 1922
415 uTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleA 432
1923 ATACAAGAGACTTCCCAAAAACACACAGACGAATTTGTATCTCTTATCA 1972
432 sncYsSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
1973 ATTGCTCTCTCAGCCACTCTGTGATCAGCCATGGTATCGGCAAGATGTG 2022
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG1 465
2023 GAGTCCTGCCATGATATGGCTGCGCTGAACATCTTAAAGTTGCTGTCTGA 2072
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
2073 GTTGGACCAACAAGTACAGAGATGCCAAGACAGAAACGACCAATGT 2122
482 eAlaLacysGlyArgCys 487
2123 CTGTGTGTGGAGGTGC 2139
```

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:293327

seq_documentation_block:

ID 293327 standard; cDNA: 3260 BP.

XX 293327;

DT 04-JUL-2000 (first entry)

DE Human stauflen cDNA.

XX Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; ; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 409..1311

FT CDS /*tag= a

FT CDS /product= Stauflen protein

XX CA2238656-A1.

XX PD 22-NOV-1999.

XX PF 22-MAY-1998; 98CA-2238656.

XX PR 22-MAY-1998; 98CA-2238656.

XX PA (UYMO-) UNIV MONTREAL.

XX PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;

XX PI Moulard AJ;

XX DR WPI: 2000-246924/22.

XX DR P-PSDB; Y83108.

XX PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful

PT for treating retroviral infections especially human immunodeficiency

PT virus infections (HIV)

XX PS Claim 4; Figure 1b; 96pp; English.

XX CC Stauflen is a RNA binding protein which interacts with double stranded

CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA

CC transport and localization. Mammalian SP contain multiple ds

CC RNA-binding domains. Chimeric proteins comprising portions of

CC stauflen with proteins with RNase or protease activity can be used to

CC treat virus infections. The RNase or protease activity of the fusion

CC protein prevents proper maturation of the virus. Mammalian stauflen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauflen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293327.

XX Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

alignment_scores:

Quality: 2316.50 Length: 489

Ratio: 4.908 Gaps: 2

Percent Similarity: 96.524 Percent Identity: 90.798

alignment_block:

US-09-316-048-11 x 293327 ..

Align seg 1/1 to: 293327 from: 1 to: 3260

```
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
430 ATGTATTAAGCCGTGTGACCCTTACTCTCGATGCAAGTCC..TATTACTA 476
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
477 CAACATGAGAGAGAGGCTTATCCCCGAGGTACTTTTACCCTATTCAG 526
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnIlePhe 50
|||||
527 TTCACACCTTACTTATCAAGTGAACCTTCTGTGGAGACAGCAATT 576
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
577 AATGCCAAAGGAAGACACAGACAGGCTGCCAACAACAGATGCTGCTCCA 626
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
|||||
627 AGCGTTGAGGATCTTCAGAAATGAGCCCTGCCAGAGAGGCTGGAGGTGA 676
84 sncGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
|||||
677 ATGGAAGAGAATCCGAAGAAATAATCTCAATTAATCTGAATAATGCA 726
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
727 GTGTTTGAGATTGCACCTTAACGGAACCTGCTGTGAATTCGAGGTGGC 776
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
|||||
777 CCGGAGAGAGTGGCCCAACCCACATGAAGACTTGTGACCAAGGTTTCGG 826
134 aGlyGluPheValGlyGluGlyGlyLysSerLysLysIleSerLys 150
|||||
827 TTGGGGAGTTTGTGGGGGAAGGTGAAGGGAAGAAACGAAGAAGATTCA 876
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProle 167
|||||
877 AAAAATGCCGCATAGCTGTCTTGAGGAGCTGAAGAAGTTACCGCCCT 926
167 uProAlaValAluArgValLysProArgIleLysLysLysSerGlnProT 184
|||||
927 GCCTGAGTTGAACAGAGTAAAGCCTAGAAATCAAAAAAGAAACAAACCA 976
184 hrcYsLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||||
977 TAGTCAAGCCACAGACAAAGCCAGAAATATGGCCAGGGGATCAATCCGATT 1026
```


199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215
1027 AGCCGACTGGCCAGATCCAGACGACAAAGAGAGAGAGAGAGAGAGTA 1076
215 rMetLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetG 232
1077 CACGCTCCTCAGAGAGAGGCTCCCGCGCCAGAGAGTTGTGATGC 1126
232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
1127 AGGTGAAGGTGGAACACACACTGCAGAGAAGAACGGGACCAACAAGAAG 1176
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 265
1177 GTGGCCAAGCGCAATGCAGCCGAGAACATATGCTGGAGATCTTGTTCAA 1226
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
1227 AGTCCCGCAGCGGAGCCACCAACCCGCACTCAAGTCAGAGAGAGAAGA 1276
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
1277 CACCCATAAAGAAACACAGGGAGTGAAGAAAGTAACCTTTTGAACCT 1326
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
1327 GGCCTCTGGGGATGAAATGGGACTAGTAATAAGAGAGATGAGTTCAGGAT 1376
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
1377 GCCTATCTAAGTCATCAGCAGCTGCCCTGCTGGAATTTCTCCCATGGTGC 1426
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
1427 CCGAGGTGCGCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGAT 1476
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetI 365
1477 TTTACCAAGGGCAGCTCCGAATCCTGCCAAGGCCACGGTAACCTGCCATGAT 1526
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382
1527 AGCCCGAGAGTTGTTGTATGGGGCACCTCGCCACAGCCGAGACCATTT 1576
382 euLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
1577 TAAAGATAACATCTCTTCAGGCCACGTAACCCCATGGACTCTTCACGAGA 1626
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGl 415
1627 CCCTCTGACCAACTGGACTATCTTTCAGAGATCCAGGGATTCCAGGTTGA 1676
415 uTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeuIleA 432
1677 ATACAAGACTTCCCAAAAACACAAGAAGCAATTTGTATCTCTTATCA 1726
432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
1727 ATTGCTCTCTCAGCCACCTCTGTATCAGCATGGTATCGCAAGATGTG 1776
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGl 465
1777 GAGTCCCTGCCATGATATGGCTGCGCTGAACATCTTAAGTTGCTGCTGA 1826
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
1827 GTTGACCAACAAGTACAGAGATGCCAAGAAGAGAAACGACCAATGT 1876
482 eAlaCysGlyArgCys 487
1877 CTGTGTGTGGAGGTGC 1893

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:X39991

seq_documentation_block:
ID X39991 standard; DNA; 1127 BP.
XX
AC X39991;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI: 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 625-626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;

alignment_scores:
Quality: 1710.00 Length: 375
Ratio: 4.763 Gaps: 2
Percent Similarity: 95.733 Percent Identity: 89.067

alignment_block:
US-09-316-048-11 x X39991 ..
Align seg 1/1 to: X39991 from: 1 to: 1127
47 GLYGlnGlnPheAsnGlyLysGlyLysMetArgProProValLysHisAs 63
|||||

3 GGACAGCAATTTAATGGCAAGGAAGACACAGACAGGCTGCGAAACACGA 52
63 pAlaProAlaArgAlaLeuArgThrLeuGlnSerGluProLeuProGluA 80
53 TGCTGCTGCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCCTGCCAGAGA 102
80 rGLeuGluValAsnGlyArgGluAlaGluGluAsnLeuAsnLysSer 96
103 GGCTGGAGGTGAATGGAGAGAATCCGAGAGAAATCTCAATAAATCT 152
97 GluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValas 113
153 GAAATAAGTCAAGTGTGAGATTGCACCTTAACGGAACCTTGCCCTGTGA 202
113 nPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheValT 130
203 TTTTCGAGGTGGCCCGGAGAGTGGCCACCCACATGAGAAACTTTGTGA 252
130 hrArgValSerValGlyGluPheValGlyGluGlyLysSerLys 146
253 CCAAGGTTTCGGTGGGAGTTTGTGGGGAAGGTGAAGGAAAGCAAG 302
147 LysIleSerLysLysAsnAlaAlaArgAlaValLeuGluGlnLeuArgAr 163
303 AAGATTTCAAGAAATAATGCCGCATAGCTGTCTTGAGAGAGCTGAAGAA 352
163 gLeuProProLeuProAlaValGluArgValLysProArgIleLysLysL 180
353 GTTACCGCCCTGCTGCAGTGAACGAGTAAGCCTAGATCAAAAGA 402
180 ySSerGlnProThrCysLys.....ThrAlaProAspTyrGlyGlnGly 194
403 AAACAATAACCCATAGTCAAGCCACAGACAGAGCCCAAGATATGGCCAGGG 452
195 MetAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLy 211
453 ATCAATCCGATTAGCCGACTGGCCACAGATCCAGCAGGCAAAAAGAGAA 502
211 sGluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgG 228
503 GGAGCCAGAGTACAGCTCTCCACAGAGCGAGCCCTCCGCGCGAGGG 552
228 LuPheValMetGlnValLysValGlyHisIsthrAlaGluGlyValGly 244
553 AGTTTGTGATGCAGGTGAAGGTTGGAACACACTGCAGAAGAACGGGC 602
245 ThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluI 261
603 ACCAACAAGAGAGGTGGCAAGCGCAATGCAGCCGAGAACATGCTGAGAT 652
261 eleuGlyPheLysValProGlnAlaGlnProAlaLysProAlaLeuLysS 278
653 CCTGGTTTCAAGTCCCGCAGGCGAGCCCAACCAACCCGACCTCAAGT 702
278 erGluGluLysThrProValLysLysProGlyAspGlyArgLysValThr 294
703 CAGAGGAGAGACACCCATAAGAAACACAGGGGATGGAAGAAAAGTAACC 752
295 phePheGluProSerProGlyAspGluAsnGlyThrSerAsnLysAspGl 311
753 TTTTGAACCTGGCTCTGGGGATGAAATGGGACTAGTAATAAGAGGA 802
311 uGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleL 328
803 TGAGTTCAAGATGCCCTATCTAAGTCAATCAGCAGCTGCTGGAATTC 852
328 euProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyHis 344
853 TTCCCATGGTGCCCGAGGTCGCCACAGGCTGAGAGAGTGAAGGACAT 902
345 HisThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrVa 361
903 CACACCAAAGATTTTACCAAGGCGAGCTCCGAATCTGCCAAGGCCACGGGT 952

361 lThrAlaMetIleAlaArgGluLeuLeuTyrGlyThrSerProThra 378
953 AACTGCCATGATAGCCCGAGAGTGTGTATGGGGGCACCTCGCCACAG 1002
378 la.GluThrIleLeuLysSerAsnIleSerSerGlyHisValProHisGl 394
1003 CCCGAGACCATTTTAAAGATAATACATCTCTTCAGGCCACGATACCCATGG 1052
394 yProArgThrArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnG 411
1053 ACCTCTCAGAGACCCCTNTGAGCAACTGCACTATCTTNCAGAGTCCAGG 1102
411 lypPheGlnValGluTyrLysAsp 418
1103 GATTNCAGGTTGAATACCAAGAC 1125
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X39992
seq_documentation_block:
ID X39992 standard; DNA; 773 BP.
XX AC X39992;
XX DT 02-JUL-1999 (first entry)
XX DE Prostate cancer associated gene.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX OS Homo sapiens.
XX PN WO9904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX DR WPI; 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX Claim 67; Page 626; 787pp; English.
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product; and (b) determining the
CC product complexed with an HLA molecule; and (c) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast

alignment_block:
US-09-316-048-11 x X40000 ..
Align seg 1/1 to: X40000 from: 1 to: 844

101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
3 GTGTTTGAGATTGCACCTTAACGGAACCTGCTGTGAATTCGAGGTGCC 52

117 aargGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
|||||
53 CCGGGAGAGTGGCCCAACCCACATGAGAACCTTGTGACCAAGGTTTCGG 102

134 aIGlyGluPheValGlyGluGlyGlyLysSerLysLysIleSerLys 150
|||||
103 TTGGGGAGTTGTGGGGGAAGGTGAAGGGAACAAGAGATTTCAAAG 152

151 LysAsnAlaAlaArgAlaValLeuGluLeuArgLeuProProLe 167
|||||
153 AAAAATGCCCGCATAGCTGTCTTGAGGAGCTGAAGAAGTTACCGCCCT 202

167 uProAlaValGluArgValLysProArgIleLysLysSerGlnProT 184
|||||
203 GCCTGCAGTTGAACGAGTAAAGCCTAGAAATCAAAAGAAACAAACCCA 252

184 hrcysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||
253 TAGTCAAGCCACAGACAGACCCAGAATATGGCCAGGGGATCAATCCGAT 302

199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTy 215
|||||
303 AGCCGACTGGCCCAAGATCCAGCAGGCAAAAGAGAGAGAGCCAGAGTA 352

215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|
353 CACGCTCCTCACAGAGCGAGCCCTCCCGCCGCGAGGAGTTGTGATGC 402

232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
403 AGGTGAAGTTGGAAACACACTGCAGAGAGACGGGCACCAACAGAAAG 452

249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 265
|||||
453 GTGGCCAAGCGCAATGCAGCCGAGAAACATGCTGGAGATCCTGTGTTCAA 502

265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
|||||
503 AGTCCCGCAGCGCGCCACCAACCCGCACTCAAGTCAGAGAGAGAAGA 552

282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|||||
553 CACCCATAAAGAAACCGGGGATGGAAGAAAGTAACCTTTTGTGAACCT 602

299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
::: |||||
603 GCTCTTGGGATGAAATGGGACTAGTAATTAAGAGGATGAGTTTCAGGAT 652

315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
|||||
653 GCCTTATCTAAGTCATCAGCAGCTGCTGCGAATTCTTTCATGCTGTC 702

332 roGluValAlaGlnAlaValGly.ValSerGlnGlyHisHisThrLysAs 348
||
703 CCGANGTCGCCCAAGCTGTAGGAAGTTAGTCAAGGACATNACACCAAGA 752

348 pPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetI 365
|||||
753 TTTTACCAGGCAGCTT...CGAATCTTGCCANGGCCNCNGTACTGCCATGA 799

365 leAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAla 378
|||||
800 TAGCCCANAGTTGTGTATTGGGGCANCCTTTGCCCCAGGCC 840

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X40002
seq_documentation_block:
ID X40002 standard; DNA; 946 BP.
XX
AC X40002;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629-630; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other;

alignment_scores:
Quality: 1121.50 Length: 312
Ratio: 4.185 Gaps: 4
Percent Similarity: 85.897 Percent Identity: 76.603

alignment_block:
US-09-316-048-11 x X40002 ..
Align seg 1/1 to: X40002 from: 1 to: 946

[illegible]

seq_documentation_block:	
ID	280269 standard; cDNA; 727 BP.
XX	
AC	280269;
XX	
DT	07-APR-2000 (first entry)
DE	Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.
XX	
KW	Human; gene expression product; diagnosis; tumour; colon cancer;
KW	colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW	hyperplasia; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO9964576-A2.
XX	
PD	16-DEC-1999.
XX	
PF	09-JUN-1999; 99WO-IB01062.
XX	
PR	10-JUN-1998; 98US-0088801.
XX	
PA	(FARB) BAYER CORP.

XX	Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI	Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI	Schlegel R;
XX	
DR	WPI; 2000-087220/07.
XX	
PT	Novel nucleic acids, used to develop products for the diagnosis and
PT	treatment of disorders involving unwanted cell proliferation,
PT	particularly cancers, especially colon cancer -
XX	
PS	Claim 15; Page 273; 469pp; English.
XX	
CC	Z79917 to Z80766 represent double stranded cDNA clones isolated from the
CC	human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC	cDNA clones can be used to generate antisense oligonucleotides which
CC	can be used for antisense therapy. Methods and products from the present
CC	invention can be used for identifying and/or classifying cancerous cells
CC	present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC	and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC	for developing agents for the diagnosis and treatment of disorders
CC	involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC	hyperplasia.
XX	
Sequence	727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;

```
alignment_scores:
  Quality: 633.00      Length: 166
  Ratio: 4.336        Gaps: 3
  Percent Similarity: 87.952      Percent Identity: 77.711

alignment_block:
```

US-09-316-048-11 x Z80269 ..
Align seg 1/1 to: Z80269 from: 1 to: 727

28	TyrPheIyrrProPheProValProProLeuLeuTyrGlnValGluLeuSe	44
3	TACTTTTACCATTTCAGTTCCAGCTTACTTTATCAAGTGGAACTTTC	52
44	rValGlyGlyGlnGlnPheasnGlyLysGlyLysMetArgProProValL	61
53	TGTGGGAGCAGACGACATTTAATGGCAAGAAGAACAGACAGAGCTGCCA	102
61	YSHISASpAlaProAlaArgAlaLeuArgThrLeuGlnSerGlnProLeu	77
103	AACACGATGCTGCTGCCAAAGCGTTGAGGATCTCTGCAGATGAGCCCTG	152
78	ProGluArgLeuGluValAsnGlyArgGluAlaGluGluLysLeuAs	94
153	CCAGAGAGGCTGGAGGTGAATGAGAGAAATCCGAAGAAGAAAATCTCAA	202
94	nlysserGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuP	111
203	TAAATCTGAATAAGTCAAGTGTTTGAGATTGCACTTAACCGAACTTGC	252
111	roValAsnPheGluValAlaAlaArgGluSerGlyProProHisMetLysAsn	127
253	CTGTGAATTTGAGGTGGCCCGGAGAGTGGCCACCCACACATGAAGAAC	302
128	PheValThrArgValSerValGlyGluPheValGlyGlyGlyGlyLy	144
303	TTTGTGACCAAGGTTTCGGTTGGGAGATTGTGGGGGAAGTGAAGGAA	352
144	sSerLysLysIleSerLysLysAsnAlaAlaArgAlaValLeuGlnGln	160
353	AAGCAAGAAGATTTCAAAGAAGAAAATGCCCGCCATAGCTGNTCTTGAGGAG	402
161	LeuArgArgLeuProProLeuProAlaVal..GluArgValLysProAr	176
403	CTGAAGAAAGTACCAGCCCTGCTGTGATTTGGACCGGAAGTTTAAGGCTTAN	452


```

277 sSerGluGluIuLysThrProValLysProGluAspGlyArgLysValT 294
676 .....GCCA 679
294 hrPhePheGluProSerProGluAspGluAsnGlyThrSerAsnLysAsp 310
    ||| ::| ||||| ||||| |||
680 AATTAGTAATATTCTCCA.....GAGAACCACTATTCT..... 714
311 GluGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyI 327
714 ..... 714
327 eleuProMetValProGluValAlaGlnAlaValGlyValSerGlnGly. 343
    ::|::|::|::|::|::|::|::|::|::|::|::|::|
715 .....TTAACAAATGTAGTAGGACATTCTTTAGGAT 745
344 .....HisHisThrLysAspPheThrArgAlaAlaProAsnProAla 357
    ||| ::|::|::|
746 GTACTGTGGCATTCCTTGAGGAAT..... 768
358 LysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyTh 374
768 ..... 768
374 rSerProThrAlaGlu..ThrIleLeuLysSerAsnIleSerSerGlyH 390
    ||||| ::|::|::|::|::|::|::|::|::|
769 .TCTCCTGTGAAAGAAGATCAACTTACTGAAAGAAGCCCTCCTTAGT... 813
390 IValProHisGlyProArgThrArgProSerGluGlnLeuTyrTyrLeu 406
    |||::|::|::|::|::|::|::|
814 .....ATTCCAATACAGATTACATCCAGCTGCTTAGTGAAATT 852
407 SerArgAlaGlnGlyPheGlnValAlaGluTyrLysAspPheProLys... 421
    ::|::|::|::|::|::|::|::|::|::|::|::|::|
853 GCCAAGCAACAGGTTTAAATATTAACATATTGTGATATAGATGAACGTAG 902
422 ...AsnAsnLysAsnGluCysValSerLeuIleAsnCysSerSerGlnP 437
    |||::|::|::|::|::|::|::|::|::|::|::|::|
903 CGCCAATGGACAATATCAATGT....CTTGCTGAACGTGCCACCAGCC 946
437 roProLeuValSerHisGlyIleGlyLysAspValGlnSerCysHisAsp 453
    ||| ||| ||||| ||| ::| ::|::|::|::|::|
947 CCATCACAGTCTGTCAATGCGTCGGTATCTCCTGTGCCAATGCACAAAGT 996
454 MetAlaAlaLeuAsnIleLeuLysLeu..... 463
    ||||| ||| |||::|::|::|::|
997 GATGCAGCTCACAAATGCTTTGCCAGTATTTAAAGATATAGCAGAAGAAA 1046
464 .....SerGluLeuAspGlnGlnSerThrGlnMetProArgThrG 477
    |||::|::|::|::|::|::|::|::|::|::|::|::|
1047 GTAAATCTGAGCAACTTAATAAATCTTTAGTAGCACAATAAAGTTCC 1096
477 LysAsnGlyPro 480
    ::|::|::|::|
1097 CCTCTGGGCCCC 1107

```

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:X03457

seq_documentation_block:

ID X03457 standard; cDNA; 1608 BP.

AC X03457;

DT 07-APR-1999 (first entry)

DE cDNA encoding a double-stranded RNA-binding protein.

KW Human; double-stranded RNA-binding protein; BINDR; Incyte clone 620438;

KW teratocarcinoma; cancer; ss.

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	CDS	115..1056
FT		/*tag= a
FT		/product= BINDR

W09851795-A1.

19-NOV-1998.

13-MAY-1998; 98WO-US09877.

13-MAY-1997; 97US-0855518.

(INCY-) INCYTE PHARM INC.

Corley NC, Hillman JL, Lal P;

WPI; 1999-059701/05.

Double-stranded RNA-binding protein - used for treating or preventing various cancers

Claim 4; Fig 1A-E; 61pp; English.

The present sequence encodes a human double-stranded RNA-binding protein (BINDR). The BINDR cDNA sequence was first identified in Incyte clone 620438 from the paranganglion cDNA library PGANNOT01. BINDR can be used in a pharmaceutical composition for treating or preventing cancer. The BINDR polynucleotide can be used for the detection of polynucleotides encoding BINDR. The cancers that can be treated using BINDR include adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma, and particularly cancers of the adrenal gland, bladder, bone, brain, breast, cervix, gall bladder, gastrointestinal tract, heart, kidney, liver, lung, ovaries, pancreas, paragangliomas, parathyroid, pituitary gland, prostate, salivary gland, spleen, stomach, thymus, thyroid, testes, and uterus.

SQ Sequence 1608 BP; 481 A; 349 C; 344 G; 434 T; 0 other;

alignment_scores:

Quality:	219.50	Length:	471
----------	--------	---------	-----

Ratio: 1.081 Gaps: 17

Similarity: 43.100 Percent Identity: 23.355

alignment_block:

16-048-11 x X03457 . . .

Align seg 1/1 to: X03457 from: 1 to: 1608

54 GlyLysMetArgProProValLysHisAspAlaProAlaArgAlaLeuAr 70

39 GGGACCAGCGACCCGGGAGCGAGCAGCAGTCTGCTCCGACCGCTCTTCCTC 88

70 gThr.....LeuGIInSerGIuProLeuProGLuArgLeuGIuValA 84

89 CAGCCGCTGAGCCGTCCTTCTCGCCATGTCCAGAGCAGGCAACCGCC 138

84 snglyArgGluAla.GluGluGluAsnLeu..... 93

139 GAGGCCCGCCGCTGGAGCGCGAGGACAGTGGGACCTTCAGTTTGGGGA 188

94AsnlyssErgluIeSerglNvalphegluI 104

189 GATGATAACAGCTAAGCCAGGGAACACACCGATTTCAGGTATTACACGAAT 238

104 LeAlaLeuLys...ArgAsnLeuProValAsnPhcIuValAlaArgGlu 119

239 ACGCATGAGACCAAGACATCCAGTT...TATGAATGTGAAGATCT 285

120 SerGlyProProHisMetLysAsnPhenValThrArgValSerValGlyGly 136

This image shows a vertical strip of a film negative. It features a series of sprocket holes along the left edge. The main body of the strip contains several horizontal lines and patterns of dots, which are likely artifacts from the scanning process or the original film's content. The overall appearance is that of a technical or archival document.

OM of: US-09-316-048-6 to: Issued_Patents_NA:* out_format : pfs
Date: Apr 5, 2001 8:31 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09316048/runat_01042001_170039_7753/app_query.fasta_1.1293
-DB=Issued_Patents_NA -QFM=fastap -SUFFIX=p2n.rn1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=biosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09316048_@CGN1_1_106 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:
Query: US-09-316-048-6
Query length: 577
Database: Issued_Patents_NA:*
Database sequences: 280836
Database length: 80580151
Search time (sec): 112.270000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-933-750C-95				218.00	376.22	2.1e-13 1904
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-234-613-95				218.00	376.22	2.1e-13 1904
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-855-518-2				211.50	366.02	7.8e-13 1608
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-280-443-1				157.00	243.15	5.4e-06 6671
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-459-1				157.00	243.15	5.4e-06 6671
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-555-678-1				157.00	243.15	5.4e-06 6671
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02275-1				157.00	205.80	0.0007 5022
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-289-368-1				135.50	175.62	0.0314 32207
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-770-379-20				132.50	203.81	0.0008 2628
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-261B-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-837-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-342-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-646A-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-076A-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-4				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-335A-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-729-214-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-028-933A-6				128.00	168.42	0.0790 28958
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-1				128.00	161.40	0.1942 49377
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-931-999-4				127.50	186.61	0.0077 6755
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-436-771-8				126.50	197.46	0.0019 2562
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-434-998-8				126.50	197.46	0.0019 2562
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-487-797-8				126.50	197.46	0.0019 2562
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02058-8				126.50	197.46	0.0019 2562
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-320-559-1				121.50	165.32	0.1175 14255
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-327-392-1				121.50	165.32	0.1175 14255
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-306-691B-55				121.50	165.32	0.1175 14255
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-545-860D-1				121.50	165.32	0.1175 14255
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-04496-1				121.50	165.32	0.1175 14255
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-961-083-3				119.50	189.37	0.0054 1714
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-1				118.50	144.98	1.60 43280
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-452-655B-1				116.00	161.57	0.1902 8532
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-450-582-1				116.00	161.57	0.1902 8532
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-790-912-1				115.50	164.81	0.1255 6201
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-741-940-1				115.50	159.05	0.2627 9606
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-289-548A-1				115.50	159.05	0.2627 9606
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-452-654-1				115.50	159.05	0.2627 9606
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-370-235A-1				115.50	159.05	0.2627 9606
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-961-083-37				113.50	180.94	0.0158 1360

/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-073-384C-2 - 112.50 171.05 0.0564 2496
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-254-359A-2 - 112.50 171.05 0.0564 2496
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-483-043-2 - 112.50 171.05 0.0564 2496
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-481-238-2 - 112.50 171.05 0.0564 2496
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-066B-2 - 112.50 171.05 0.0564 2496

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-933-750C-95

seq_documentation_block:

: Sequence 95, Application US/08933750C
: Patent No. 5932442
: GENERAL INFORMATION:
: APPLICANT: Lal, Preeti
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Bandman, Olga
: APPLICANT: Shah, Purvi
: APPLICANT: Au-Young, Janice
: APPLICANT: Yue, Henry
: APPLICANT: Guegler, Karl J.
: APPLICANT: Corley, Neil C.
: TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
: NUMBER OF SEQUENCES: 98
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: FILING DATE: September 23, 1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0356 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 95:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1904 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: THP1A2S08
: CLONE: 2754573
: US-08-933-750C-95

alignment_scores:

Quality: 218.00 Length: 490
Ratio: 1.03 Gaps: 15
Percent Similarity: 43.061 Percent Identity: 22.041

alignment_block:

US-09-316-048-6 x US-08-933-750C-95 ..

Align seg 1/1 to: US-08-933-750C-95 from: 1 to: 1904

134 GYGlyGInGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHI 150
||||| ||| ::::: ||

```
17 GGTGGCGCGCTGGANNTCCCTCCGGGGACCAGCGACCCGGGAGCATGCA 66
150 SASpAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuPro. 166
    :::||| :::|||::: ::: |||||
67 CGTCGCGCACCACTTCACGTAGGCTTGGAAACAGACAACTAACTCCCTC 116
166 ..... 166
117 CATCTCCAACCTGCTTGACAAAGCTCGCTCTCCTCCAGCCGCTGAGCCG 166
167 .....GluArgLeuGluValAsnGlyAr 174
    :::||| :::|||:::
167 TCCCTTCTCGCCATGTCCAGAGCAGCGACCGCGCCGAGCCCGCGCT 216
174 gGluSerGluGluGluAsnLeu..... 181
    ||| |||:::
217 GGAGCGCGAGCAGAGTGGGACCTTCAGTTTGGGGAAGATGATACAGCTA 266
182 ....AsnLysSerGluIleSerGlnValPheGluIleAlaLeuLys... 195
    :::|||::: |||::: ||| ::::
267 AGCCAGGGAACACACCGATTTCAGGTATACACGAATACGGCATGAAGACC 316
196 ArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProH1 212
    :::|||::: |||::: ||| ::::
317 AAGACATCCCACTT...TATGAATGTGAAGATCTGATGTGCAAAATACA 363
212 smetLysAsnPheValThrLysValSerValGlyGluPheValGlyGluG 229
    :::|||::: :::|||::: ||| ::::
364 CGTCCCACTTTCACCTTCAGAGTAACTGTTGGTGACATTAACCTGCACAG 413
229 IyGluGlyLysSerLysLysIleSerLysLysAsnAlaAlaAlaVal 245
    ||||| |||||::: ||||| ::::
414 GTGAAGGTACAAGTAAAGAGCTGGCAACATAGAGCTGCAGAGCTGCC 463
246 LeuGluGluLeuLysLysLeuProProLeuProAlaValGluArgVally 262
    :::: ||||| ::::
464 ATAAACATTTTGA...GCAA 483
262 sProArgIleLysLysLysThrLysProIleValLysProGlnThrSerP 279
    :||| ::: |||::: |||
484 TGCAAGTATTTGCTTGCAGTTCCTGACCCCTTAATGCCCTGACCCCTTCCA 533
279 roGluTyrgLynGlyIleAsnProIleSerArgLeuAlaGlnIleGln 295
    ::: ::: |||||::: ||| ::::
534 AGCAACCAAGAACCACTTAATCTATTCCTATTCATTACAGAAATTGGCT 583
296 GlnAlaLysLysGluLysGluProGluTyrThrLeuLeuThrGluArgG1 312
    ::: |||||::: |||
584 ATTCAATCATGGCTGGAGACTTCCTGAATATACCTTTCCAGAGGAGGAGG 633
312 yLeuProArgArgArgGluPheValMetGlnValLysValGlyAsnHisT 329
    :::: |||||::: ::::
634 ACCTGCTCATAGAGAGATATATCTACATTTGACAGGCTAGAGTCATTTTA 683
329 hrAlaGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaAla 345
    ||| |||::: ||||| ||||| |||||
684 TGGAAACTGGAAAGGGGCATCAAAAAAGCAAGCCAAAGGAATGCTGCT 733
346 GluAsnMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProTh 362
    |||::: |||
734 GAGAAATTTCTT..... 745
362 rLysProAlaLeuLysSerGluGluLysThrProIleLysLysProGlyA 379
745 ..... 745
379 spGlyArgLysValThrPhePheGluProGlySerGlyAspGluAsnGly 395
    ||| ::: |||::: |||||
746 ....GCCAAATTTAGTATATTTCTCCA.....GAGAACAC 778
396 ThrSerAsnLysGluAspGluPheArgMetProTyrlLeuSerHisGlnG1 412
    |||
779 ATTTCT..... 784
```

```
412 nLeuProAlaGlyIleLeuPrometValProGluValAlaGlnAlaValG 429
785 .....TTACAAATGTAGTAG 800
429 IyValSerGlyGly.....HisHisThrLysAspPheThrArgAla 442
    || ||| ||| ::::
801 GACATTTCTTAGATGATGACTTGGCATTCCTTGAGGAAT..... 838
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459
838 ..... 838
459 uLeuTyrgLynGlyThrSerProThrAlaGlu...ThrIleLeuLysAsnA 475
    ||||| ::: :::: |||||:::
839 .....TCTCCTGGTGAAGATCAACTTACTGAAAGAA 872
475 snIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 491
    :::: ||| :::|||::: |||
873 GCCTCCTTAGT....ATCCAAAT.....ACAGATTACATCCAG 907
492 GlnLeuAspTyrlLeuSerArgValGlnGlyPheGlnValGluTyrlLysAs 508
    ||||| :::: |||||::: |||
908 CTGCTTAGTGAATTTGCCAAGAACAGGTTTATATATACATATTGGA 957
508 pPheProLysAsnAsnLysAsn...GluPheValSerLeuIleAsnCysS 524
    ::: ::: ||| :::: |||
958 TATAGATGAAGTGAAGCGCCCAATGACATATATCAATGTCTTGCTGAACCTGT 1007
524 erSerGlnProProLeuIleSerHisGlyIleGlyLysAspValGluSer 540
    ||::: ||| ::: ||||| ||| ::::
1008 CCACGAGCCCAATCACAGTCTGTCAATGCTCCGATATCTCCTGTGGCAAT 1057
541 CysHisAspMetAlaAlaLeuAsnIleLeuLysLeu..... 553
    :::: ||||| ||| |||::: |||
1058 GCACAAGTGAATGCAGCTCACAATGCTTGCAGTATTAAAGATAATAGC 1107
554 .....SerGluLeuAspGlnGlnSerThrGluMetP 564
    |||::: ||| ::: |||:::
1108 AGAAGAAGATAATCTGAGCAACTTAATAATCTTTCAGTAGCACATA 1157
564 roArgThrGlyAsnGlyPro 570
    ::: ::: |||||
1158 AAAAGTTCCTCCCTCTGCCCC 1177
seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-234-613-95
seq_documentation_block:
; Sequence 95, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
```



```

1892 AATTCAGAGAATCATCCACTATTCACAGAGAAAGATCAGAGAGAC 1941
272 eVallySPROGLInThr.....SerProGLuTYrGLyGLInGLyI 285
      ::::: ||||| ::::: |||||
1942 TGCAGAGTCCAGACCCACCCTTCAGCCACATCCTTCTTTCTGGGA 1991
285 leasnProIleSerArgLeuAlaGLInIleGLInAlaLysLysGLuLys 301
      ::|||::: ||| ::::: |||
1992 AGAGCCCCGTCACACACTGCTGTAGTGTATGCAC...AATTTGGGGAAC 2038
302 GLuProGLuTYrThrLeuLeuThrGLuArgGLyLeuProArgArgArgL 318
      :::: |||::: |||||::: |||
2039 TCCTGCCAATTCCGTCCTCTGTCCAAAGAGGCCCTGCCCATGAACCAA 2088
318 upheValMetGLInValLysValGLyAsnHisThrAlaGLuGLy...Thrg 334
      ::||| ||| ||||| ::||| ::::: |||
2089 GTTCCAATACTGTGTGTGAGTGGAGCCCAACTTTCCCAAGTGTAGTG 2138
334 lYThrAsnLysLysValAlaLysArgAsnAlaAlaGLuAsnMetLeuGLu 350
      :::: ||||| |||||::: |||||::: |||||
2139 CTCCAGACGAAGAGTGCAGAAAGCAGATGGCCGACAGAGAGCCATGAAG 2188
351 IleLeu.....GLyPheLysValProGLInArgGLInPro... 361
      ||| ::::: |||
2189 GCCCTGCATGGGAGGCGACCAACTCCATGGCTTCTGTATACCAAGCCCTGA 2238
362 .....ThrLysProAlaLeuLysSerGLuGLuLysThrP 373
      :::: |||||
2239 AGGTATGATCTCAGAGTCACTTGATACCTTGGAAATCCATGATGCCCAACA 2288
373 roIleLysLysProGLyAspGLyArgLysValThrPhePheGLuProGLy 389
      :::: ||| ||||| :::: |||
2289 AGGTCAGGAAGATTGGCAGCTGTGAGATACCTGAACACCAACCCTGTG 2338
390 SerGLy..... 391
      ::|||
2339 GGTGGCCTTTTGGAGTAGCGCCGCTCCATGGCTTGTGCTGTAATTCAA 2388
392 .....AspGLuAsnGLyThrSerAsnLysGLuAspGLuPheArgMetP 406
      |||::: |||
2389 GTTGTGTCAGACAGTCCGAGCTCTCTACAGAG.....C 2420
406 roTYrLeuSerHisGLInLeuProAlaGLy...IleLeuProMetVal 421
      || :::: |||||
2421 CCAAGTTCGTTTACCAAGCAAAAGTTGGGGTTCGCTGTTCCACAGCCGTC 2470
422 .....ProGLuValAlaGLInAla..... 427
      ::|||
2471 TGGCGACACAGCAAGCAAGGCAAGCAGAGCAAGCAGATGCGGCTCT 2520
428 .....ValGLyValSerGLInGLyHisHisThrLysAspPheThrA 441
      ::||| ::::: |||||
2521 CCGTGTCTTGTATGGGAGAGACGAGAAGGCAAGCATGGGTTTCACAG 2570
441 rgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 457
      :::: |||
2571 AGGTAACCCA.....GTGACAGGGGCCAGTCTCAGAGA 2605
458 GluLeuLeuTYrGLyGLyThrSerProThrAlaGLuThrIleLeuLysAs 474
      ::||| :::: |||||
2606 ACTATGCTCTCTCTCAAGGTCCCCAGAGCAGAG..... 2641
474 nasnIleSerSerGLyHisValProHisGLyProLeuThrArgPro.... 489
      ||| ||| |||
2642 .....CCAAAGACACTCCCTCTCA 2660
490 .....SerGLuGLInLeuAspTYrLeuSerArgValGLInGLy 501
      ::|||::: |||||
2661 CTGGCAGACACCTTCCATGACACAGATAGCCATGCTGAGCCAC..... 2701
502 PheGLInValGLuTYrLysAspPheProLysAsnAsnLysAsnGLuPheVa 518
      :::: |||
2702 .....CGGTGCTTCAA 2712

```

```

518 lSerLeuIleAsnCysSerSerGLInProProLeuIleSerHisGLyIleG 535
      ::||| ||| ||| ||||| ||::: |||
2713 CACTCTGACTAAC...AGCTTCAGACCCCTCTGCTCGGCCGACAGATT. 2758
535 lYlYsAspValGLuSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 551
      ::||| |||::: |||
2759 .....CTGGCCGCATC...ATTATGAAA 2779
552 LeuLeuSerGLuLeuAspGLInSerThrGLuMetProArgThrGLyAs 568
      ||||| ||::: |||
2780 AAAGACTCTGAG...GACATGGGTGTGTCGTCAAGCTTGGAAACAGGAA 2826
568 nGLYPrometSerValCysGLyArgCys 577
      | |||||
2827 T.....CGCTGT 2833
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-459-1
seq_documentation_block:
; Sequence 1, Application US/08457459
; Patent No. 5677428
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,459
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49CUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 155..3832
; US-08-457-459-1

```

```

alignment_scores:
Quality: 157.00      Length: 626
Ratio: 0.553        Gaps: 31
Percent Similarity: 45.367      Percent Identity: 22.364

```



```

535 lyllyaspvalglusercysHisaspmetalaaleuasnlleulys 551
      ::|||::|
2759 .....CTGCCGCCATC..ATTATGAAA 2779
      |||||
552 leuleusergluleuaspglInserThrglumeTProargThrglyas 568
      |||||
2780 AAAGACTCTGAG...GACATGGGTGTCGTCTGACCTTGGAACAGGAA 2826
      |||||
568 nglyPrometserValcysglyargCys 577
      |||||
2827 T.....CGCTGT 2833

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-555-678-1

```

seq_documentation_block:
: Sequence 1, Application US/08555678
: Patent No. 5763174
: GENERAL INFORMATION:
: APPLICANT: Nishikura, Kazuko
: TITLE OF INVENTION: RNA Editing Enzyme and Methods
: NUMBER OF SEQUENCES: 67
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Howson and Howson
: STREET: Spring House Corporate Cntr, P.O. Box 457
: CITY: Spring House
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19477
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/555,678
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/197,794
: FILING DATE: 17-FEB-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/280,443
: FILING DATE: 25-JUL-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,459
: FILING DATE: 01-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Bak, Mary E.
: REGISTRATION NUMBER: 31,215
: REFERENCE/DOCKET NUMBER: WST49DUSA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-540-9206
: TELEFAX: 215-540-5818
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6671 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 155..3832
: US-08-555-678-1

```

alignment_scores: Quality: 157.00 length: 626
 Ratio: 0.553 Gaps: 31
 Percent Similarity: 45.367 Percent Identity: 22.364

alignment_block:

US-09-316-048-6 x US-08-555-678-1 ..
 Align seg 1/1 to: US-08-555-678-1 from: 1 to: 6671

```

33 MetSerIleProSerThrThrSerSerLeuProSerGluAsnAlaGlyAr 49
      |||::||| ::|||::||| |||::|||
1241 ATGCAATCAAGAGAAATACGAACAGTGTCTCT...GAACCGCTCCAGC 1287
      |||
49 gProIle.....GlnAsnSerAlaLeuProSerAlaSerIleT 62
      |||
1288 TGCAATCCCTGAGACCAAGAAAGAACGACAGATTCTCACCCTGTAATATAC 1337
      |||
62 hrSerThrSerAlaAlaAlaGluSerIleThrProThrValGluLeuAsn 78
      ::|||::|||::| ::|
1338 CCACATCAATGCGCTCAATAATACATGTTAACCCACAGAAAGTGCAGCAT 1387
      |||
79 AlaLeuCysMetLysLeuGlyLysLysProMetTyrLysProValAspPr 95
      |||::|||::|||::|
1388 .....GGGCAGGAACCTGTCTATAAG.....TT 1410
      |||
95 OTyrSerArgMetGluSerThrTyrAsnTyrAsnMetArgGlyAlaAT 112
      ::|||::|||::|||
1411 AGAAACAGCAAGAGAGCCAGA..... 1432
      |||
112 YrProProArgTyrPheTyrProPheProVal.....ProProLeu 125
      |||
1433 .....CCAGAACCAAGCAAGACTGAACACCACCTGT 1462
      |||
126 LeuTyrGlnValGluLeuSerValGlyGlyGlnInPhe..AsnGlyLy 141
      |||::|
1463 CATTACAATGGCCCTCAAAAGACAGGATATGTGACTTTGAAAATGGCCA 1512
      |||
141 sglyLysThr.....ArgGlnAla 148
      |||
1513 GTGGGCCACAGATGACATCCAGATGACTTGAATAGTATCCGGCAGCAC 1562
      |||
148 lalysHisaspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluPro 164
      ::|
1563 CAGGTGAGTTTCGACCATCATGAGATGCCCTCTTACACATCATGGC 1612
      |||
165 LeuProGluArgLeuGluValAsnGlyArgGluSerGluGluGluAsnLe 181
      |||
1613 TTGCCA.....CGGTCTCACCTACAAAGAACT 1641
      |||
181 uasn.....LysSerGluIleSerGlnValPheGluIleAla 194
      |||
1642 GACAGAGTGCACGCTGAGAAAGAACCCCATCATCAGCGGCTGTAGAAATGCC 1691
      |||
194 eulysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 210
      ::|
1692 AGTTCCGCTAGTCAAACTGTGAGTTCACACATGATAGACAGAGTGCACCA 1741
      |||
211 ProHisMetLysAsnPheValThrLysValSerVal..GlyGluPheVa 226
      |||||
1742 CCCCATGAACCTCGATTTAATTCACAGTTGTTCATCATGGCCGACAGTT 1791
      |||
226 lGlyGluGlyGluGlyLysSerLysLysIleSerLysLysAsnAlaAla 243
      ::|||::|||::|||::|||::|||::|||
1792 TCCCCCAGCTGAAGCTGGAAGCAAGAAAGTGCCCAAGCAGATGCAGCTA 1841
      |||
243 le.....AlaValLeuGluGluLeuLysLysLeuProProLeu 255
      ::|
1842 TGAAGCCATGACAAATTCTGCTAGAGGAAGCAAAAGCCAAAGACAGTGA 1891
      |||
256 ProAlaValGluArgValLysProArgIleLysLysLysThrLysProIle 272
      ::|
1892 AAATCAGAGAATCATCCCACTATTCCACAGAGAAAGAAATCAGAGAGAC 1941
      |||
272 eValLysProGlnThr.....SerProGluTyrGlyGlnGlyI 285
      ::|||::|||::|||
1942 TGCAGAGTCCACAGACCCCAACCTTTCAGCCACATCCTTCTTTCTGGGA 1991
      |||
285 leasnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLys 301

```


1241 ATGCAATCAAGAAATACGACAGTGTCT...GAAACCGCTCCAGC 1287
49 gProIle.....GlnAsnSerAlaLeuProSerAlaSerIleT 62
1288 TGCATTCCTGAGACCAAAAGAAACGAGAGTTCCTCACCCTGTAATATAC 1337
62 hrSerThrSerAlaAlaIaIaGluSerIleThrProThrValGluLeuAsn 78
1338 CCACATCAATGCCCTCAATAATACATGTGTACACAGAAAGAGTGAGAAAT 1387
79 AlaLeuCysMetLysLeuGlyLysLysProMetTyrLysProValAspPr 95
1388GGGCAAGAACCTGTCAATAAG.....TT 1410
95 oTyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArgGlyGlyAlaT 112
1411 AGAAACACAGCAAGAGCCACA..... 1432
112 yrProProArgTyrPheTyrProPheProVal.....ProProLeu 125
1433CCAGAACCCAGCAAGACTGAAACCACCTGT 1462
126 LeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe...AsnGlyLys 141
1463 CATTAACAATGCCCTCAAAAGCAGGATGTGTACTTTGAAAAATGGCCA 1512
141 sGlyLysThr.....ArgGlnAla 148
1513 GTGGGCCACAGATGACATCCAGATGACTTGAAATAGTATCCGGCCAGCAC 1562
148 lAlYSHisAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluPro 164
1563 CAGGTGAGTTTCGAGCCATCATGGAGATGCCCTCTCTACAGTCAATGCC 1612
165 LeuProGluArgLeuGlnValAsnGlyArgGluSerGluGluGluAsnLe 181
1613 TTGCCA.....CGGTGTTCAACCCTACAAAGAACT 1641
181 uAsn.....LysSerGluIleSerGlnValPheGluIleAla 194
1642 GACAGAGTGCAGCTGAAGAACCACCATCAGCGGCTGTAGAAATATGCC 1691
194 eulYsArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 210
1692 AGTTCGCTAGTCAAACTGTGAGTTCACACATGATAGACAGAGTGACCA 1741
211 ProHisMetLysAsnPheValThrLysValSerVal...GlyGluPheVa 226
1742 CCCCATGAACCTCGATTTAAATTCAGGTTCATCAATGCGCCGAGACTT 1791
226 lGlyGluGlyGlyLysSerLysLysIleSerLysLysAsnAlaAlaI 243
1792 TCCCCCAGCTGAAGCTGAAGCAAGAAAGTGCCCAAGCAGATGCAGCTA 1841
243 le.....AlaValLeuGluGluLeuLysLysLeuProProLeu 255
1842 TGAAGCCCATGACAATTCTGTAGAGGAAGCCAAAGCCAAAGCAGAGTGA 1891
256 ProAlaValGluArgValLysProArgIleLysLysLysThrLysProI 272
1892 AAATCAGAGAATCATCCCACTATTCCACAGAGAAAGAAATCAGAGAGAC 1941
272 eValLysProGlnThr.....SerProGluTyrGlyGlnGlyI 285
1942 TGCAGAGTCCAGACCCCAACCCCTTCAGCCACATCCTCTTTTCGGGA 1991
285 leAsnProIleSerArgLeuAlaGlnIleGlnAlaLysLysGluLys 301
1992 AGAGCCCCGTACACACACTGCTTGAGTGTATGCAC...AAATTGGGAAC 2038
302 GluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgG 318
2039 TCCTGCGAATTCCGTCTCTGTCCAAAGAGGCCCTGCCCATGAACCCAA 2088

318 upheValMetGlnValLysValGlyAsnHisThrAlaGluGly...Thrg 334
2089 GTTCCAATACTGTGTGGCAGTGGAGGCCCAACTTCCCCAGTGTGAGTG 2138
334 lYThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 350
2139 CTCCAGCAAGAAAGTGGCAAAAGCAGATGGCCGAGAGGAAGCATGAAG 2188
351 lLeu.....GlyPheLysValProGlnArgGlnPro.. 361
2189 GCCCTGCATGGGAGCGCAGCAACTCCATGGCTTGTGATAACCAAGCTGA 2238
362ThrLysProAlaLeuLysSerGluLysThrP 373
2239 AGGTATGATCTCAGAGTCACTTGATACTGGAAATCCATGATGCCCAACA 2288
373 roIleLysLysProGlyAspGlyArgLysValThrPhePheGluProGly 389
2289 AGGTCAGGAAGATTGGCGAGCTCGTGAATACCTGAACACCAACCTGTG 2338
390 SerGly..... 391
2339 GGTGGCCTTTGGAGTAGCCGCCCTCCCATGGCTTGTGCTGTAATTCAA 2388
392AspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetP 406
2389 GTTGTGCAGCAGTCCGAGCTCCTCAGAG.....C 2420
406 roTyrLeuSerHisGlnGlnLeuProAlaGly...lLeuProMetVal 421
2421 CCAAGTTCGTTTACCAAGCAAAAGTTGGGGTTCGCTGTCCAGCCGTC 2470
422ProGluValAlaGlnAla..... 427
2471 TGGCCACACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTGGGCTCTP 2520
428ValGlyValSerGlnGlyHisHisThrLysAspPheThra 441
2521 CCGTGTCTTGATTTGGGAGAGAAGCAAGCAAGCAAGCATGGGTTTCACAG 2570
441 rGAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 457
2571 AGGTAACCCA.....GTGACAGGGGCCAGTCTCAGAAGA 2605
458 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAs 474
2606 ACTATGCTCTCTCTCAAGGTCCCCAGAACACAG..... 2641
474 nasnIleSerSerGlyHisValProHisGlyProLeuThrArgPro.... 489
2642CCAAAGACACTCCCTCTCA 2660
490SerGluGlnLeuAspTyrLeuSerArgValGlnGly 501
2661 CTGGCAGCACCTTCATGACCAGATAGCCATGCTGAGCCAC..... 2701
502 pheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheVa 518
2702CGGTGCTTCAA 2712
518 lSerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyIleG 535
2713 CACTGTGACTAAC...AGCTTCAGGCCCTCTGTGCTGGCCGCAAGATT. 2758
535 lYlYsAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 551
2759CTGGCCGCCATC...ATTATGAAA 2779
552 leuLeuSerGluLeuAspGlnGlnSerThrLumMetProArgThrGlyAs 568
2780 AAAGACTCTGAG...GACATGGGTGTCTGTCAGCTTGGAACAGGGAA 2826


```

2125 ...AGTGGACAGGAATGACATCACCACCCACCGACAGAG... 2166
402 lupheargmetProTyrLeuSerHisGlnGlnLeuProAlaGlyLeu 418
2167 .....CTC 2169
419 PrometValProGlu.....ValAlaGlnAlaVala 428
2170 CCACCAAGCCCAAGAAAGCAAGCCTTTGGCCACCACTCAACCTGC 2219
428 lGlyValSerGlnGlyHisHisThrLysAspPheThrArgAlaAlaProA 445
2220 AAAGACTTCAMCATCGAAAGCCAAAGACAGCCCACTTCTCCCTTAAGC 2269
445 snProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyr 461
2270 AGCCAGCTCCCAACCAACCATTTGGTGGTGAATAAAACCCTATG... 2313
462 GlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnIleSerSe 478
2314 .....AGCCTTGCTTC 2324
478 rGlyHisValProHisGlyProLeuThrArgProSer 490
2325 AGGCTTAGTGCCAGCTGCCCAACCAAGCCCTGCC 2361

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-770-379-20

seq_documentation_block:
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

```

```

alignment_scores:
    Quality: 132.50      Length: 496
    Ratio: 0.616        Gaps: 24
    Percent Similarity: 43.347      Percent Identity: 20.161

alignment_block:
US-09-316-048-6 x US-08-770-379-20 ..
Align seg 1/1 to: US-08-770-379-20 from: 1 to: 32207

36 ProSerThrThrSerSer.....LeuProSerGluAsnAla..... 47
4882 CCAGATAGTGAATCTCTCTCTGAGCTGTGGCCAGATCGGAGCCATTGC 4931
48 .....GlyArgProIleG 52
4932 CCTCCCTGGCCGCCGACACGAGAGCGGTGCCACTGGGAGACCCGTGC 4981
52 lnsSerAlaLeuProSerAlaSerIleThrSerThrSerAlaAla 68
4982 GCCACATTCACACCCCGA.....AGAGATACAGATCTTTCGCGC 5022
69 GluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysLeuG 85
5023 TCAAGCCGCTGCTTCAATATACCTTCGCAATTA..... 5058
85 yLysLysProMetTyrLysProValasp.ProTyrSerArgMetGlnSer 101
5059 .CCAGATCCCAACGACACGACAAACCCGATACCGATGTTGTGCTAAACGCT 5107
102 ThrTyrAsnTyrAsnMetArgGly.....Al 111
5108 .....AACAAATAACCTTGAACAAACAGCTACATCCCTCGCGATCGCAAGC 5151
111 aTyrProArgTyrPheTyrProPheProValProProLeuLeuTyrG 128
5152 GGACCGCTACGACCATCTATGTAGTCCCTCTGAAGCG..... 5191
128 lnsValGluLeuSerValGlyGlnGlnPheAsnGlyLysGlyThr 144
5192 .....CAGGATAGATGG 5203
145 ArgGln.....AlaAlaLys 149
5204 CCTGAATAATGACCAACGACAAACCCCTGCGGCTACCGACATCGCC 5253
149 shisAspAlaAlaLysAlaLeuArgIleLeuGlnAsnGlu..... 163
5254 AGAAGAGCGGGATCACAGTTGCCAGACTCCTTCAGAGCCAGGTGAGCG 5303
164 .....ProLeuProGluArgLeuValAsn 172
5304 CCACATGCTTGACATCTTCTCCAGGTTCCGAGCGCTCGCTCCCAAT 5353
173 GlyArgGluSerGluGluGluAsnLeuAsnLysSerGluIleSerGlnVa 189
5354 GCGCCTCAGAGAGTTCCAGACATG.....TCAGAAAT 5388
189 lPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAla 206
5389 CCACGAGGTGCC.....GCCACGCGCCACTCACAATACCCCAATA 5432
206 rgluSerGlyProProHisMetLysAsnPheValThrLysValSerVal 222
5433 AACGACCGGAAACCCCTCACGTCTCCCGGAGAGGTGATCAATAACAGAA 5482
223 GlylupheValGlyGluGlyLysSerLysLysLysSerLysLys 239
5483 CGCAAA.....CGCGACAGCAGCCGAGATTTGCGCGGA 5517
239 sasnaAlaAlaIleAlaValLeuGluLeuLysLysLeuProLeuP 256

```



```

418 .....GAAAGAAGGCAGTTAGCTTATTATTATGACACAA 454
277 hrSerPro.....GluTyrGlyGlnGlyIleAsnProIleSer 289
||::: ||::: ||::: ||:::
455 CGAATTCTTCAGAAGCATTTATCCATGGGAAATTACATAGCCTTATCAAT 504
290 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluProGluTyrTh 306
||::: ||::: ||::: ||:::
505 AGAATTGCCACG.....AAGAAAGACTAACTGTAATTTATGA 542
306 rLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGln 323
::: ||::: ||::: ||:::
543 ACAGTGTGCATCG...GGGGTCATGGGCCAGAGGATTTTCATTATTAAT 589
323 allysValGly..AsnHisThrAlaGluGlyThrGlyThrAsnLys 338
||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
590 GCAAAATGGGACAGAAAGAATATAGTATTGGTACAGGTTCTACTAAACAG 639
339 ValAlaLysArgAsnAlaAlaGlu..AsnMetLeuGluIleLeuGlyPh 354
||||::: ||::: ||::: ||::: ||::: ||:::
640 GAAGCAAAACAATTTGGCCCGCTAAACTTGCATATCTTCAGATATTA.... 684
354 elysValProGlnArgGlnProThrLysProAlaLeuLysSerGluIul 371
||||::: ||::: ||::: ||::: ||::: ||:::
685 .....TCAGAGAGAA. 693
371 ySThrProIleLysLysProGlyAspGlyArgLysValThrPhePheGlu 387
693 ..... 693
388 ProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspGluPheAr 404
|||| ||::: ||::: ||:::
694 .....ACCTCAGTGAATCTGAC..... 711
404 gMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetV 421
|||| ||::: ||::: ||::: ||::: ||:::
712 .....TACCTGTCC..... 720
421 alProGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisHisThrLys 437
::: ||::: ||::: ||::: ||::: ||:::
721 .....TCTGGT 726
438 AspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMe 454
::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
727 TCTTTTGCTACTACGTGTGAGTCCCAAGCAACTCTTTAGTGACCACAC 776
454 tIleAlaArgGluLeuLeuTyrGly....ThrSerProThrAlaG 469
::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
777 ACTGCTTCTGAATCATCATCTGAAGGTGACTTCTCAGCAGATACATCAG 826
469 luthrIleLeuLysAsnAsnIleSerSer 478
|| ||::: ||::: ||::: ||::: ||::: ||::: ||:::
827 AGATAATCTACAGTGACAGTTTAAACAGT 855

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-261B-6
seq_documentation_block:
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF INVENTION: antipathogenic substances
; CORRESPONDENCE ADDRESS:
;

```

```

: ADDRESS: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/258,261B
: FILING DATE: 08-JUN-1994
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
:
US-08-258-261B-6

```

[illegible]

```
10839 .....AGGCCACTCGC.....CCGCTGCTCTC 10863
125 euleuTyrgluValgluleuSerValglYglngluInpheasnlyLys 141
      |||||
10864 TCCTG.....GGCGAGCTCTCTAC..... 10883
142 gLyLsthrargGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuAr 158
      |||
10884 .....CCGAGGCTCGCGCTCGACTGGAAGACTTCTTCGCGCCCTACG 10927
158 gileuGln.....AsnluProleuProgluArgLeu... 169
      |||
10928 CTCCCCGCAAGTCTCCCTCCACCCTACCCCTTCCAGCAGAGCGGTTTC 10977
170 .....GluValAsnlyArg..... 174
      |||
10978 TGGCTCGACGTCTCCACGAGCAGAACGCTTCGACGTCGCCCTCCGACGCC 11027
175 .....GluSerGluGluGluAsnleuAsnly 183
11028 TGACCTCGGCGCCACCAATCCCGCTGCTGGCGCGCGCTGCGCTTCGCG 11077
183 sSerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnleuProV 200
      :
11078 ACCGCGGTGGCTTCTCTTACAGAGCGGCTCTCCCTCGCAGACACCCG 11127
200 al..... 200
11128 TGGCTCGAAGGCCATGCCGTCTTCGGACACACCATTCTACGGGACCCG 11177
201 .....AsnPh 202
11178 CTTTCTCGAGCTCGCCCTGCACGTGCGCCACCGCGTCGGCTCGACACCG 11227
202 egluValAlaArggluSerGly.Pro.....ProHisMetLysasn 215
      :
11228 TCGAAGAGCTCACGCTCGAGGCCCTCTGCTCTCCATCGCAGACACC 11277
216 pheValThrLysValSerVal.GlygluPheValGlygluGlyLyl 232
      :
11278 GTCCTCTCCAGATCTCCGCGCGCTGAGCAGCAGCAGCAGCAGAGGCG 11327
232 ysSerLysLysIleSerLysLysAsn..... 240
      :
11328 GCTCTCTTCCATAGCGGACACAAGAGAGCGGCTTCAGATGCCCCCTGA 11377
241 ..AlaAlaIleAlaValleuGluGluLeuLysLysLeuProProleuPr 256
      :
11378 CTCGCCACGCCAGCGCTCTCTCGCGCGGACCCCATCTCTCCGCC 11427
256 alaValgluArgValLysProArgIleLysLysLysThrLysProIleV 273
      :
11428 GATCTCCACGAGTGGCTCCCTCGAGTCCATCCGCGTGACCTCGAAGG 11477
273 alLysProGlnThrSerProgluTyrgluGlynglyIleasnProIleSer 289
      :
11478 CCTCTACGCAACCTCGCAACCTCGGGCTTGCTTACGGCCCGAGTTCC 11527
290 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysgluProgluTyrrh 306
      :
11528 AGGGCCTCC..... 11536
306 rleuLeuThrGluArgGlyLeuProArgArgGluPheValMetGlnV 323
11536 ..... 11536
323 alLysValglYasnHisThrAlaGluGlyThrGlyThrAsnLysLysVal 339
      :
11537 .....GCTCCGCTTACAGAGCGCGGCGCAGAGCTCTTTGCGGAGCCA 11578
340 AlalysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysVa 356
      :
11579 AGCTCCCGG.....AAGCG 11592
```

```
356 lProGlnArgGlnProThrLysProAlaLeuLysSerGluGluLysThr. 372
      :
11593 GCCGAAAAGGATGCCG.....CCCGTTTGCCCTCCACCCCTGCGCTGCT 11636
373 .....ProIleLysLysProGlyAspGlyArg 381
      :
11637 CGACAGCGCCCTGCATGACTGGCCTTTGAGAGACGACAGAGAGGACGG 11686
382 LysValThrPhePheGluProGlySerGlyAspGluAsnlyThrSerAs 398
      :
11687 TCGCTCTGC.....CCTTCTGTGAGCGGAGTCTCGCTGCGCT.. 11725
398 nLysGluAspGluPheArgMetProTyrrLeuSerHisGlnGlnLeuProA 415
11725 ..... 11725
415 lAglyIleLeuPrometValPro.GluValAlaGlnAlaValGlyValSe 431
      :
11726 .....CCGTGCGTGCACCACTTGCCTGCGCTTCACCGCTC 11764
431 rGlnGly..... 433
11765 CCAAGGGTGAATCTCCGCTCTCGATGCTCTGCGCGGACGCCGAGGTGAC 11814
434 .....HisIsthrLysAspPheThrArgAla 442
11815 CTTCTTGCTCGGTGCAAGCGCTCGCATGCGGAGCAGCTCCGCGCGCA 11864
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459
      :
11865 GCTCCGACACCC..... 11876
459 uLeuTyrglyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnI 476
      :
11877 .....GGCAGCTTCCCA...CCATGATGCGGCTCTTCGCGCTGAC 11913
476 lSerSerGlyHisValProHis.....GlyProLeuThrArg 488
      :
11914 TGGAGCGAGCTCCAAAGCCCACTTCACCGGCTCGCGCCGAGGCGGT 11963
489 ProSerGluGlnLeuAspTyrrLeuSerArg.ValGlnGlyPheGlnValG 505
      :
11964 CTTCTCGGACAGCGGCGGACGATCTCGCGCTGACGCGCCGCTGCGCC 12013
505 luTyrrLysAspPheProLysAsnAsnLysAsnGluPheValSerLeuIle 521
      :
12014 GCTACGCCGACCTCGCTGCCCTCCGAAGC.....GCCCTCGAC 12051
522 AsnCySerSerGlnPro.....ProLeuIleSerHisgl 533
      :
12052 CAGGCGGCTTCGCTCCGCGCTCGTCTGCGCCCTTCATCGATCGAGCC 12101
533 yIleGlyLysAspValgluSerCysHisAspMetAlaAlaLeuAsnIleL 550
      :
12102 GGCAGGCGACCTCGTCCCGAGCGCCACGAGGCCACCGCGCTCGACTGC 12151
550 euLysLeu 552
12152 CCTCTTTG 12159
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-837-6
seq_documentation_block:
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
```

APPLICANT: Lam, Stephen Ting
 APPLICANT: Hammer, Phillip E.
 APPLICANT: Uknes, Scott Joseph
 TITLE OF INVENTION: Genes for the synthesis of
 TITLE OF INVENTION: antipathogenic substances
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/456,837
 FILING DATE: 01-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/457,205
 FILING DATE: 01-JUN-1995
 APPLICATION NUMBER: 08/258,261
 FILING DATE: 08-Jun-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8614
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-456-837-6

alignment_scores:
 Quality: 128.00 Length: 656
 Ratio: 0.489 Gaps: 30
 Percent Similarity: 39.939 Percent Identity: 20.427

alignment_block:

US-09-316-048-6 x US-08-456-837-6 ..

Align seg 1/1 to: US-08-456-837-6 from: 1 to: 28958

18 SerGlnIleuAsnLysAsnGlnSerLeuSerGlnProLeu..... 32
 :::::::::::::::::::: ::::::::::::::::::::
 10574 ACATCGCTCTCGGACGTGCGAGCTCCTTTATTGACCGTCACCGGC 10623
 33MetSerIleProSerThrThrSerSerLeuProSerGluAsn 47
 :::::::::::::::::::: ::::::::::::::::::::
 10624 ACCAGGCTGACGCGCTCCGAGCTCGACGCGCGTACTGTATCGAAACCT 10673
 47 lagIArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSer 63
 :::::::::::::::::::: ::::::::::::::::::::
 10674 CCGCAACCGTCTCTGTCGAGCGCGACCGAGCGGCTCTCGACGATG 10723
 64 ThrSerAlaIleAlaGluSerIleThrProThrValGluLeuAsnAlaLe 80
 :::::::::::::::::::: ::::::::::::::::::::
 10724 GGCATCGCTCTCTCGGTCGAGGTACGCCCC..... 10753
 80 uCysMetLysLeuGlyLysLysProMetTyrLysProValAsp..... 94

10754ATCCCGTGTACAGCTCGCCCTCCGCGAGACCTGCGAGCGCTCACC 10799
 95ProTyrSer.ArgMetGlnSerThrTyrAsnTyrAsnMetAr 108
 10800 GCTGCATCCCGTCTCGTCTCGGCTCCATTCGACGAGAGA..... 10838
 108 gLgLyAlaIleProProArgTyrPheTyrProPheProValProProL 125
 10839AGGCCACCTCGC.....CCGCTGCTCTCCTC 10863
 125 eULeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLys 141
 10864 TCCTG.....GGCGAGCTCTCTAC..... 10883
 142 GlyLysThrArgGlnAlaAlaLysHisAspAlaAlaLysAlaLeuAr 158
 10884CCGAGGCTCTCGCTCGACTGGAAGACTTCTTCGCGCCCTACG 10927
 158 gIleLeuGln.....AsnGluProLeuProGluArgLeu.... 169
 10928 CTCCCGGCAAGGTCTCCCTCCCGCACCTACCCCTTCACGAGAGCGGTTC 10977
 170GluValAsnGlyArg..... 174
 10978 TGGCTGACGCTCTCCACGAGCAGAACGCTTCCGACGTGCGCTCCGACAGGCC 11027
 175GluSerGluGluGluAsnLeuAsnLys 183
 11028 TGACCTGGCGGCAACCAATCCCGTGTGCGCGCGCGCTTCGCGCG 11077
 183 sSerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProV 200
 11078 ACCGCGGTGGCTTCTCTTACAGGGCGGCTTCCCTCGACAGACACCGG 11127
 200 aL..... 200
 11128 TGGCTCGAAGGCCATGCCGTCTTCGGCACACCATCTACCGGGCACCGG 11177
 201AsnPh 202
 11178 CTTTCTGACGCTCGCCCTGCACGTCGCCACCGCGTGGCGCTGCACACCG 11227
 202 eGluValAlaArgGluSerGly.Pro.....ProHisMetLysAsn 215
 11228 TCGAAGAGCTCACGCTCGAGGCCCTCTCGCTCTCCATCGCAGACACCC 11277
 216 PheValThrLysValSerVal.GlyGluPheValGlyGlyGlyGlyL 232
 11278 GTCTCTCTCCAGATCTCTCGTGGGCGCGTGGACGACGACGAGCAAGGGC 11327
 232 ySerLysLysIleSerLysLysAsn..... 240
 11328 GCTCTCTTCCATAGCCGACAGAGAGCGGCTTCAGATGGCCCTCGA 11377
 241 ..AlaAlaIleAlaValLeuGluGluLeuLysLysLeuProProLeuPr 256
 11378 CTCGCCACGCCAGCGGCTCTCTCGCGCGGCGACCCCATCCCTCTCGGCC 11427
 256 oAlaValGluArgValLysProArgIleLysLysLysThrLysProIleV 273
 11428 GATCTCCACGAGTGGCTCTCCCTCGAGTGCATCCCGGTGACCTCGAAGG 11477
 273 aLysProGlnThrSerProGluTyrGlyGlnGlyIleAsnProIleSer 289
 11478 CCTCTACGCAACCTCGCAACCTCGGGCTTGCCTACGGCGCCGAGTTCC 11527
 290 ArgLeuAlaGlnIleGlnAlaLysLysGluLysGluProGluTyrTh 306
 11528 AGGCGCTCC..... 11536
 306 rLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGlnV 323

```

11536 ..... 11536
323 allyValGlyAsnHisThrAlaGlyThrGlyThrAsnLysLysVal 339
      ::::: |||::: ::::: |||
11537 .....GCTCCGCTTACAAGCGCGGCGAGCAGCTCTTTGCCGAAGCCA 11578
340 AlaLysArgAsnAlaAlaGluAsnMetLeuGluLeuGlyPheLysVa 356
      ::::: |||
11579 AGCTCCCGG.....AAGCG 11592
356 lProGlnArgGlnProThrLysProAlaLeuLysSerGluLysThr. 372
      |||::: |||::: |||::: |||:::
11593 GCCGAAAAGGATGCCG.....CCCGTTTGGCTCCACCCCTGCGCTGCT 11636
373 .....ProLeLysLysProGlyAspGlyArg 381
      |||::: |||::: |||::: |||:::
11637 CGACAGCGCCCTGCATGCACTGGCCTTTGAGGACGACAGAGAGGACGG 11686
382 LysValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAs 398
      ::::: |||::: |||::: |||::: |||:::
11687 TCGCTCTGC.....CCTTCTCGTGAGCGGAGTCTCGCTGCGCT. 11725
398 nLysGluAspGluPheArgMetProTyrLeuSerHisGlnLeuProA 415
11725 ..... 11725
415 laGlyIleLeuPrometValPro.GluValAlaGlnAlaValGlyValSe 431
      ||| ||||| ::::: |||::: |||::: |||
11726 .....CCGTGCGGCCACACCACTTGGCGCTTCCACCGCTC 11764
431 rGlnGly..... 433
      |||||
11765 CCAAGGGTGAATCCTCCGCTCGATCGTCCGCGGACGCCGAGGTGAC 11814
434 .....HisHisThrLysAspPheThrArgAla 442
      ::::: |||::: |||::: |||::: |||:::
11815 CCTCTTGCCCTCGGTGCAAGCGCTCGCATGCGGAGCAGCTCCGCCGCA 11864
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459
      |||||::: |||
11865 GCTCCGCACCCC..... 11876
459 uLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnI 476
      |||::: |||::: |||::: |||:::
11877 .....GGCAGCTTCCCA...CCATGATGCGCTCTCCGCGTGCAC 11913
476 lSerSerGlyHisValProHis.....GlyProLeuThrArg 488
      ::::: |||::: |||::: |||::: |||:::
11914 TGAAGCGAGCTCCAAAGCCCACTTACCGCGCTGCCGCCGAGCGCGCT 11963
489 ProSerGluGlnLeuAspTyrLeuSerArg.ValGlnGlyPheGlnValG 505
      |||||::: |||::: |||::: |||:::
11964 CCTTCTCGGACAGGCGGCGCAGATCTCGCGCTCGACGCCCGCTCGCCC 12013
505 lutyrlYAspPheProLysAsnLysAsnGluPheValSerLeuIle 521
      ::||| |||::: ::::: |||
12014 GCTACGCCGACCTCGCTGCCCTCCGAAGC.....GCCCTGCAC 12051
522 AsnCySerSerGlnPro.....ProLeuIleSerHisG1 533
      ::::: |||::: |||::: |||::: |||:::
12052 CAGGGCGCTTCGCGCTCCGCGCTCGTGCCTGCCCTTCATGATGCAC 12101
533 ylieGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleL 550
      ||| ||| |||::: |||::: |||::: |||:::
12102 GGCAGGCGACCTCGTCCGAGCGGCCACAGAGGCCACCGCGCTCGACTCG 12151
550 euLysLeu 552
      |||
12152 CCCTCTTG 12159
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-342-6
seq_documentation_block:

```

```

; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-342-6

alignment_scores:
  Quality: 128.00      Length: 656
  Ratio: 0.489         Gaps: 30
  Percent Similarity: 39.939  Percent Identity: 20.427

alignment_block:
  US-09-316-048-6 x US-08-457-342-6 ..
  Align seg 1/1 to: US-08-457-342-6 from: 1 to: 28958

18 SerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeu..... 32
   ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
10574 ACATCGCTCTCGGACGTGCGAGCTCCCTCTTATTTCGACCGTCACCGGC 10623
33 .....MetSerIleProSerThrThrSerSerLeuProSerGluAsnA 47
   :: ||| |||::: |||::: |||::: |||::: |||::: |||::: |||:::
10624 ACCAGGCTCGACGGCTCCGAGCTCGACGCGCGCTACTGTATCGAAACCT 10673

```

```

47 1aGlyArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSer 63
   ::::: ||| ||||| |||
10674 CCGCAACCGCTCTGTCGAGCGGACGAGCGCTCTCGACGATG 10723
   ::::: ||| ||||| ::::: |||
64 ThrSerAlaAlaGlnSerIleThrProThrValGluLeuAsnAlaLe 80
   ||::: |||
10724 GGCATCGCTTCTCCGTCGAGTCAGCCCCC 10753
   ::::: |||
80 uCysMetLysLeuGlyLysLysProMetTyrLysProValAsp 94
   ::::: |||
10754 ...ATCCCGTGTCAAGCTCGCCCTCCGCGAGACCTGCGAGCGCTCAC 10799
   ::::: ||| ||| ||| ::::: |||
95 ...ProTyrSer.ArgMetGlnSerThrTyrAsnTyrAsnMetLar 108
   ||| ||| ||| ::::: |||
10800 GCTCGATCCGTCGTCTCGCTCCATTCGACGAGAAGA 10838
   |||
108 gGlyGlyAlaTyrProProArgTyrPheTyrProPheProValProProL 125
   ||||| |||
10839 ...AGCCCACTCGC 10863
   ||||| |||
125 euleuTyrGlnValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLys 141
   ||||| |||
10864 TCCTG 10883
   |||||
142 GlyLysThrArgGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuAr 158
   ||| ||| ||| ::::: |||
10884 ...CCGAGGCTCGCGCTCGACTGGAAGGACTTCTTCGCGCCCTACG 10927
   |||
158 gIleLeuGln 169
   ||| ||| ||| ||| ||| |||
10928 CTCCCGCAAGGTCTCTCCCTCCACCTACCCCTTCCAGCGAGCGGTTTC 10977
   |||
170 ...GluValAsnGlyArg 174
   ::::: |||||
10978 TGGCTCGACGCTCTCCAGGACGACGACGCTTCGCGCTCCGACAGGCC 11027
   |||
175 ...GluSerGlnGluGluAsnLeuAsnLys 183
   ::::: |||||
11028 TGACCTCGCGCCGACCAATCCCGCTGCTCGGCGCGCGCTCGCCTTCGCG 11077
   |||
183 sSerGlnLysSerGlnValPheGlnLysAlaLeuLysArgAsnLeuProY 200
   ::::: ||||| ||| ||| |||
11078 ACCGCGGTGGCTTCTCTTACAGGGCGGCTCTCCCTCGCAGACGACCCG 11127
   |||
200 aL 200
   |||
11128 TGGCTCGAAGGCCATGCCCTCTTCGGACACCCATCTACCGGGCACCCG 11177
   |||
201 ...AsnPh 202
   ::::: |||
11178 CTTTCTCGAGCTCGCCCTGACAGCTCGCCACCGCGTGGCCTCGACACGG 11227
   |||
202 eGluValAlaArgGlnSerGly.Pro 215
   ::::: ||||| ||| ||| ||| |||
11228 TCGAAGAGCTCACGCTCGAGGCCCTCTCGCTCTCCATCGCAGACACC 11277
   |||
216 pheValThrLysValSerVal.GlyGluPheValGlyGluGlyLysL 232
   ::::: ||||| ||| ||| ||| |||
11278 GTCTCTCTCCAGATCTCCGCGCGCGCGTGGACGACGACGAGGAGG 11327
   |||
232 ySerLysLysLysSerLysLysAsn 240
   ::||| ::||| |||
11328 GCTCTCTTTCATAGCCGACAGAGAGACGCGCTTCAGGATGGCCCTGGA 11377
   |||
241 ...AlaAlaIleAlaValLeuGluGluLeuLysLysLeuProProLeuPr 256
   ||::: ||::: ||::: ||::: ||::: |||
11378 CTCGCCACGCGCAGCGCTCTCTCGCGCGGACCCCATCTCTCCGCC 11427
   |||
256 oAlaValGluArgValLysProArgIleLysLysLysThrLysProIleV 273
   ||||| |||
11428 GATCTCCACGAGTGGCTCTCCCTCGAGTGCATCCGCGTGAGACCTCGAAG 11477
   |||
273 alLysProGlnThrSerProGluTyrGlyGlnGlyIleAsnProIleSer 289

```

```

   ::::: ||| ||||| ||| |||
11478 CCTTACGCAACCTCGCCCAACCTCGGGCTTCCTACGGCCCGAGTTCC 11527
   |||
290 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrTh 306
   ||| :::::
11528 AGGCGCTCC 11536
   |||
306 rLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetGlnV 323
   |||
11536 ... 11536
   |||
323 alLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLysVal 339
   ::::: ||||| ::::: |||
11537 ...GTCCTCTACAAAGCGCGGCGAGAGCTCTTGGCGAAGCCA 11578
   |||
340 AlalysArgAsnAlaAlaGlnAsnMetLeuGluIleLeuGlyPheLysVa 356
   ::::: |||
11579 AGCTCCCGG 11592
   |||
356 lProGlnArgGlnProThrLysProAlaLeuLysSerGlnGluLysThr 372
   ||::: ||::: ||| |||
11593 GCCGAAAGGATGCCG 11636
   |||
373 ...ProIleLysLysProGlyAspGlyArg 381
   ||| ||| ||| ||| ||| |||
11637 CGACAGCGCCCTGCATGCACCTGCGCTTGAGAGCAGACAGAGAGGAGCG 11686
   |||
382 LysValThrPheGlnProGlySerGlyAspGluAsnGlyThrSerAs 398
   ::::: ||| ||| ||| |||
11687 TCGCTCTGC 11725
   |||
398 nLysGluAspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProA 415
   |||
11725 ... 11725
   |||
415 1aGlyIleLeuProMetValPro.GluValAlaGlnAlaValGlyValSe 431
   ||| ||||| ::::: ||||| |||
11726 ...CCGTGGTGCCACCACTTGGCGGTCCACCGTC 11764
   |||
431 rGlnGly 433
   |||||
11765 CCAAGGGTGAATCTCGCTCTCGATGCTGCTGCGCCGACGCGCAGGTGAC 11814
   |||
434 ...HisHisThrLysAspPheThrArgAla 442
   ::::: ||||| ||| ||| |||
11815 CCTTGTGCTCGGTGCAAGCGCTGCGCATGCGGAGCAGCTCGCGCGCGCA 11864
   |||
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459
   ||||| |||
11865 GCTCCGCAACCC 11876
   |||
459 uLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnI 476
   ||| ||| |||
11877 ...GGCAGCTTCCCA..CCATGATGCGCTCTTCGCGGTGAC 11913
   |||
476 lSerSerGlyHisValProHis 488
   ::::: ||||| ||| ||| |||
11914 TGGAGCGAGCTCCAAGGCCCACTTACCCGCTGCGCGCGCGCGGT 11963
   |||
489 ProSerGlnGlnLeuAspTyrLeuSerArg.ValGlnGlyPheGlnValG 505
   ||||| ||| ||| ||| |||
11964 CTTTCTCGGACAGCGCGCCACGATCTCGCGCTGACGCGCGCTCGGCC 12013
   |||
505 lTyrLysAspPheProLysAsnAsnLysAsnGluPheValSerLeuIle 521
   ::||| ||| |||
12014 GTTACGCGCAGCTCGTCCCTCCGGAAGC 12051
   |||
522 AsnCysSerSerGlnPro 533
   ::::: ||| ||| ||| ||| |||
12052 CAGGGCGCTTCCCTCCGCGCTCGTGTGCGCCCTTCATCGATCGAC 12101
   |||
533 yIleGlyLysAspValGlnSerCysHisAspMetAlaAlaLeuAsnIleL 550
   ||| ||| ||| ||| ||| |||

```


12102 GGCAGGCGACCTGTCGCCGAGCGCCACGAGGCCACCGCGCTCGACATCG 12151

550 eulysleu 552

12152 CCCTCTTG 12159

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-646A-6

seq_documentation_block:

; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,646A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994

ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-646A-6

alignment_scores:
; Quality: 128.00 Length: 656
; Ratio: 0.489 Gaps: 30
; Percent Similarity: 39.939 Percent Identity: 20.427

alignment_block:
US-09-316-048-6 x US-08-457-646A-6 ..

Align seg 1/1 to: US-08-457-646A-6 from: 1 to: 28958

18 SerGlnIleuAsnLysAsnGlnSerLeuLeuSerGlnProLeu..... 32
:::|||||::: |||::: |||
10574 ACATCGCTCTCGGACGTGCGAGCTCCCTCTTATTCGACCGTCACCGGC 10623
33MetSerIleProSerThrThrSerSerLeuProSerGluAsnA 47
::: |||||::: |||::: |||:::
10624 ACCAGGCTCGACGGCTCCGAGCTCGACGGCGGCTACTGTATCGAAACCT 10673
47 laGlyArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSer 63
:::|||||::: ||| |||||::: |||
10674 CCGGCAACCGTCTGTTCTGAGCGCGACCGAGCGGCTCTCGACGATG 10723
64 ThrSerAlaAlaIleGluSerIleThrProThrValGluLeuAsnAlaIle 80
|||::: |||::: |||
10724 GGATCGCTTCTCCGTCGAGGTGAGCCCC..... 10753
80 uCysMetLysLeuGlyLysLysProMetTyrLysProValAsp..... 94
::: |||::: |||::: |||:::
10754ATCCCGTGTCTACGCTGCCCTCCGCGAGACCTGCGAGCGCTCACC 10799
95ProTyrSer.ArgMetGlnSerThrTyrAsnTyrAsnMetAr 108
||| ||| |||::: |||::: |||
10800 GCTGATCCCGCTCGTCTCGGCTCCATTCGACGAGAGA..... 10838
108 gGlyGlyAlaTyrProProArgTyrPheTyrProPheProValProProL 125
||||| ||| ||| |||
10839AGGCACCTCGC.....CCGCTGCTCTC 10863
125 euleuTyrglnValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLys 141
||||| ||||| |||::: |||
10864 TCCTG.....GGCGAGCTCTCTAC..... 10883
142 GlyLysThrArgGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuAr 158
||| |||::: |||::: ||| ||| |||
10884CCGAGGCTCGCGCTGACTGGAAGACTTCTTCGCGCCCTACG 10927
158 gIleLeuGln.....AsnGluProLeuProGluArgLeu.... 169
||| ||| ||| ||| ||| |||
10928 CTCCCGCAAGGTCTCCCTCCACCCTTCCAGCGAGAGCGGTTTC 10977
170GluValAsnGlyArg..... 174
||| ||| ||| ||| ||| |||
10978 TGGCTGACGCTCTCCAGGAGACGCTTCCGACGTGCTCCGACGCC 11027
175GluSerGluGluAsnLeuAsnLy 183
||| ||| ||| ||| ||| |||
11028 TGACCTGCGCCGACCAATCCGCTGCTGCGCGCCGCTGCGCTTCCGCCG 11077
183 sSerGluLeuSerGlnValPheGluIleAlaLeuLysArgAsnLeuProV 200
::: |||::: ||| ||| ||| ||| |||
11078 ACCGCGTGGCTTCTCTTACAGGGCGGCTCTCCCTCGACAGACCCG 11127
200 al..... 200
||| ||| ||| ||| ||| |||
11128 TGGCTCGAAGCCATGCCGTTCGGCACACCCATCTACGGGGACCCG 11177
201AsnPh 202
||| ||| ||| ||| ||| |||
11178 CTTTCTCGAGCTCGCCCTGACGCTGCCCCACCGCGCTCGGCCCTCGACACCG 11227
202 eGluValAlaArgGluSerGly.Pro.....ProHisMetLysAsn 215
||| ||| ||| ||| ||| |||
11228 TCGAAGACTCACGCTCGAGGCCCTCTCGCTCTCCATCGACGACACCG 11277
216 PheValThrLysValSerVal.GlyGluPheValGlyGluGlyLyl 232
::: |||::: ||| ||| ||| ||| |||
11278 GTCTCTCTCAGATCTCGTGGGCCCTGAGACGACGACGACGAAGGCG 11327
232 ysSerLysLysIleSerLysAsn..... 240
||| ||| ||| ||| ||| |||
11328 GCTCTCTTCCATAGCCGACAGAGACGCGCTTACGATGCCCCCTGA 11377

alignment_scores:

Quality: 128.00 Length: 656
Ratio: 0.489 Gaps: 30
Percent Similarity: 39.939 Percent Identity: 20.427

alignment_block:

US-09-316-048-6 x US-08-458-076A-6 ..

Align seg 1/1 to: US-08-458-076A-6 from: 1 to: 28958

```
18 SerGlnIleLeuAsnLysAnGlnSerLeuSerGlnProLeu..... 32
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10574 ACATCGCTCCTCGACGCTCGAGCTCCCTCTTATTGACCGTCACCGGC 10623
33 .....MetSerIleProSerThrThrSerSerLeuProSerGluAsn 47
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10624 ACCAGGCTCGACGGCTCCGAGCTCGACGGCGGTACTGGTATCGAACC 10673
47 IagIYArGProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSer 63
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10674 CCGGCAAAACCGTCTGTCTCGAGCGCGACCGAGCGGCTCTCGACGATG 10723
64 ThrSerAlaAlaIagIleSerIleThrProThrValGluLeuAsnAlaIe 80
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10724 GGCATCGCTTCTCCGTGAGGTCAAGCCCC..... 10753
80 uCysMetLysLeuGlyLysProMetTyrLysProValAsp..... 94
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10754 ....ATCCCGTGTACAGCTCGCCCTCCGCGAGACCTCGAGCGCTCACC 10799
95 .....ProTyrSer.ArgMetGlnSerThrTyrAsnTyrAsnMetAr 108
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10800 GCTCGATCCCGTCTGTCGTGCGCTCCATTCGACGAGAAGA..... 10838
108 gGlyGlyAlaTyrProProArgTyrPheTyrProPheProValProProL 125
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10839 .....AGGCCACTCTGC.....CCGCTGTCTCTC 10863
125 euLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLys 141
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10864 TCCTG.....GGCGAGCTCTCTAC..... 10883
142 GlyLysThrArgGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuAr 158
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10884 .....CCGAGGCTCTGCGCTCGACTGGAAGACTTCTCGCGCCCTACG 10927
158 gIleLeuGln.....AsnGluProLeuProGluArgLeu... 169
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10928 CTCCCGCAAGTCTCCCTCCCACTACCCCTTCCAGCGAGAGCGGTTC 10977
170 .....GluValAsnGlyArg..... 174
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
10978 TGGCTCGACGTTCTCCACGAGCAACGCTTCCGAGCTCGCTCCGAGGCC 11027
175 .....GluSerGluGluGluAsnLeuAsnLys 183
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11028 TGACCTCGGCGGACCAATCCGCTGCTCGGCGCGCGCTCGCTTCGCGG 11077
183 sSerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProv 200
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11078 ACCGGGTGGCTTCTCTTACAGGGCGGCTCTCCCTCGCAGACACCCG 11127
200 aL..... 200
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11128 TGGCTCGAAGGCATGCGTCTTCGGACACACCAATCCTACGGGACCGG 11177
201 .....AsnPh 202
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11178 CTTCCTCGAGCTCGCCCTGCACGTGCGCCACCGGTCGGCTCGACACCC 11227
202 eGluValAlaArgGluSerGly.Pro.....ProHisMetLysAsn 215
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11228 TCGAAGAGCTCACGCTCGAGGCCCCCTCTCGCTCTCCCATCGCAGACACC 11277
```

```
216 PheValThrLysValSerVal.GlyGluPheValGlyGluGlyL 232
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11278 GTCTCTCTCCAGATCTCCGTGGGCCCGGTGAGGAGCAGACGAAGGCG 11327
232 ySerLysLysIleSerLysLysAsn..... 240
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11328 GCTCTCTTCCATAGCCGACAGAGACGCGCTTCAGAGATGGCCCCCTGA 11377
241 ..AlaAlaIleAlaValLeuGluGluLeuLysLysLeuProProLeuPr 256
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11378 CTGCGCAGCGCAGCGGCTCTCTCTCGCGCGGACCCCATCTCTCGCGC 11427
256 oAlaValGluArgValLysProArgIleLysLysLysThrLysProIleV 273
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11428 GATCTCCACGAGTGGCTCCCTCGAGTGCCATCCGCTGAGACCTCGAAG 11477
273 aLysProGlnThrSerProGluTyrGlyGlnGlyIleAsnProIleSer 289
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11478 CCTCTACGCAACCTCGCCACCTCGGGCTTGCCCTACGGCCCCGAGTTCC 11527
290 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrTh 306
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11528 AGGCGCTTC..... 11536
306 rLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGlnV 323
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11536 ..... 11536
323 aLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLysVal 339
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11537 .....GCTCGCTTACACGCGCGGCGAGAGAGCTTTGCCGAAGCCA 11578
340 AlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysVa 356
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11579 AGCTCCCGG.....AAGCG 11592
356 lProGlnArgGlnProThrLysProAlaLeuLysSerGluGluLysThr. 372
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11593 GCCGAAGAAGATGCCG.....CCGGTTTGGCCCTCCACCTCGCTGCT 11636
373 .....ProIleLysLysProGlyAspGlyArg 381
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11637 CGACAGCGCCCTGCATGCACTGGCCTTTGAGAGCAGCAGAGAGGACGG 11686
382 LysValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAs 398
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11687 TCGCTCTGC.....CCTTCTGTGAGCGAGTCTCGCTGCGCT.. 11725
398 nLysGluAspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProA 415
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11725 ..... 11725
415 IagIleLeuProMetValPro.GluValAlaGlnAlaValGlyValSe 431
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11726 .....CCGTGCTGCCACCACTTGCCTGCGCTTCACACCGTTC 11764
431 rGlnGly..... 433
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11765 CCAAGGTGAATCTCTCCGTCTGCATCTCTCGGCCGAGCGCCAGGTGAC 11814
434 .....HisIsthrLysAspPheThrArgAla 442
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11815 CCTCTTGCTCGGTGCAAGCGCTCGCATGCGGAGCAGCTCCGCGCGCA 11864
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11865 GCTCCGACCCC..... 11876
459 uLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnI 476
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
11877 .....GGCAGCTTCCCA...CATGATGCGGCTCTCCGCGTCGAC 11913
```

```
476 leSerSerGlyHisValProHis.....GlyProLeuThrArg 488
      :::: :::: ||||| ||| |||
11914 TGGAGCGAGCTCCAAAGCCCCACTTTCACCGCCTGCCGCGCCGAGCGGCGT 11963
      ||||| ::::: ||||| ::::: ::::
489 ProSerGluGlnLeuAspTyrLeuSerArg.ValGlnGlyPheGlnValG 505
      ||||| ::::: ||||| ::::: ::::
11964 CCTTCTCGGCACAGCGCGCCACGATCTCGCGCTCGACGCGCCGCTCGCCC 12013
      ::||| ||::: ::::: ::|||
505 luTyrLysAspPheProLysAsnAsnLysAsnGluPheValSerLeuIle 521
      ::||| ||::: ::::: ::|||
12014 GCTACGCGGACCTCGCTGCCCTCCGAAGC.....GCCCTCGAC 12051
      522 AsnCysSerSerGlnPro.....ProLeuIleSerHisG1 533
      ::: ::||| ||| ||| ::||| :::::
12052 CAGGGCGCTTCGCCCTCCCGGCTCGTCTCGCCCCCTTCATCGATCGACC 12101
      533 yIleGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleL 550
      ||| ||| ||| ::||| :::
12102 GGCAGGCGACCTCGTCCCGAGCGCCACGAGGCCACCGCGCTCGCACTCG 12151
      550 euLysLeu 552
      |||
12152 CCCTCCTTG 12159
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:30:32 ; Search time 11371 Seconds
(without alignments)
1760.655 Million cell updates/sec

Title: US-09-316-048-9
Perfect score: 2857
Sequence: 1 cgggcgcgccccctcccccg.....ccaataatgtaataactaa 2857

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estro14:*
121: em_estro15:*
122: em_estro16:*
123: em_estro17:*
124: em_estro18:*
125: em_estro19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: gb_gss7:*
159: gb_gss8:*
160: gb_gss9:*
161: em_gss5:*
162: em_gss6:*
163: em_gss7:*
164: em_gss8:*
165: em_gss9:*
166: em_gss10:*
167: em_gss11:*
168: gb_gss10:*
169: gb_gss11:*
170: em_gss12:*
171: gb_gss12:*
172: gb_gss13:*
173: gb_gss14:*
174: gb_gss15:*
175: gb_gss16:*
176: gb_gss17:*
177: gb_gss18:*
178: gb_gss19:*
179: em_gss13:*
180: gb_gss20:*
181: gb_gss21:*
182: gb_gss22:*
183: gb_gss23:*
184: gb_gss24:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	665	23.3	688	92	AW549911	AW549911 L0057C11-
2	596.6	20.9	633	89	AW320695	AW320695 uc022b10.y
3	582.6	20.4	1018	136	BE867390	BE867390 601442470
4	554.6	19.4	725	106	BE306047	BE306047 601101723
5	554.4	19.4	624	110	BE623404	BE623404 ut91f11.x
C 6	529.8	18.5	677	92	AW555746	AW555746 L0259A06-
C 7	526.6	18.4	686	91	AW537709	AW537709 G0120F09-
8	525.8	18.4	564	2	AA120063	AA120063 mp88b08.r
9	517.2	18.1	522	139	BF016149	BF016149 uy31e08.y
10	511	17.9	522	2	AA122533	AA122533 mp95c04.r
11	508.4	17.8	752	137	BE894295	BE894295 601437666
C 12	507.8	17.8	511	92	AW542549	AW542549 C0158H10-
13	502.2	17.6	616	111	BE674157	BE674157 7d76a04.x
14	499.4	17.5	520	147	W89457	W89457 mf73e03.r1
C 15	486	17.0	497	92	AW541590	AW541590 C0146H10-
16	467	16.3	467	106	BE286356	BE286356 601097825
17	463.4	16.2	609	136	BE810330	BE810330 MR4-PT005
18	462.4	16.2	504	40	AW141293	AW141293 EST291334
19	441.2	15.4	481	4	AA240112	AA240112 my22a02.r
20	440	15.4	452	105	BE198783	BE198783 ug82e04.y
C 21	435.2	15.2	594	40	AW162002	AW162002 au72c03.x
C 22	435	15.2	528	28	AU021763	AU021763 AU021763
23	434	15.2	686	110	BE613972	BE613972 601504093
24	426	14.9	464	18	A1286843	A1286843 uc55d11.r
25	426	14.9	1801	146	U69197	U69197 U69197 Soar
26	422	14.8	724	97	AW952360	AW952360 EST364430
27	415.8	14.6	442	13	AA880492	AA880492 vx40h06.r
28	414.8	14.5	852	137	BE872137	BE872137 601446238
C 29	414.6	14.5	829	106	BE300155	BE300155 600944602
C 30	414.2	14.5	564	40	AW163206	AW163206 au97f05.x
31	406.2	14.2	446	136	BE865252	BE865252 UI-M-BH0-
32	401	14.0	401	2	AA106767	AA106767 m186f12.r
C 33	398.8	14.0	615	134	BE082712	BE082712 RC2-BT064
34	395.6	13.8	424	16	A1155343	A1155343 ud88c03.r
35	395	13.8	425	16	A1152490	A1152490 ud83e09.r
36	387.2	13.6	673	135	BE781223	BE781223 601469159
37	386.8	13.5	573	3	AA206573	AA206573 zq51f08.r
38	386	13.5	398	147	W99833	W99833 mg26f09.r1
39	384.4	13.5	524	96	AW877286	AW877286 MR4-PT005
40	379	13.3	384	2	AA124161	AA124161 mp94g04.r
41	377	13.2	377	2	AA073987	AA073987 mm95d04.r
42	376.6	13.2	391	29	AV027862	AV027862 AV027862
43	376.2	13.2	866	136	BE868328	BE868328 601443887
44	373.8	13.1	475	4	AA270608	AA270608 va66b02.r
45	372.6	13.0	543	134	BE089861	BE089861 RC5-BT070

ALIGNMENTS

RESULT 1
AW549911/c 688 bp mRNA EST 31-AUG-2000
LOCUS L0057C11-3 Mouse E12.5 Female Mesonephros and Gonads cDNA Library
DEFINITION Mus musculus CDNA clone L0057C11 3', mRNA sequence.
ACCESSION AW549911
VERSION AW549911.1 GI:7195339
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 688)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Dol,H., Wood,W.H., III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE 20381348
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: L0057 row: C column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 688
POLYA=Yes.
FEATURES
source
1. 688
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L0057C11"
/clone_lib="Mouse E12.5 Female Mesonephros and Gonads cDNA library"
/sex="female"
/dev_stage="12.5dpc"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology); Site_1: SalI; Site_2: NotI; Total RNAs were extracted from 2 Mesonephros. The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dT) primer [NotI primer-adaptor from GibcoBRL] [5'-pgactagttctagatccgcgacggccgcccctttttttttttt-3'] from 3.42ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker lI-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang."

BASE COUNT 209 a 144 c 162 g 173 t
ORIGIN

Query Match 23.3%; Score 665; DB 92; Length 688;
Best Local Similarity 99.7%; Pred. NO. 1.le-173;
Matches 687; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 2168 cacatgactgactcagtcgcgcacacgtagcacacgtgaaccactgtctctctc-a 2226
|||||
Db 688 CACATGACTGACTCAG-TCCGCCACACGTAGCACCACTGTAACCACTGCTTCTTCA 630
|||||

QY 2227 gttcatttttttccttgatgatatacaactataatttcatttcagttccttagtcg 2286
|||||
Db 629 GTTTCATTTTTCCTTGATGATACACACTATAATTTCATTTCAGTTCCTTAGTCG 570
|||||

QY 2287 tgtctacttacctagcaggttagaactgtcagtcagttaactggaagatcacagccc 2346
|||||
Db 569 TGCTCTACTTACCTAGCAGTTAGAAACTGTCACTCATGTAACCTGGCAAGATCACAGCCC 510
|||||

QY 2347 ggttggttgagcatctgtgcctctggtctggtgctgaacagttctggaattaccaccagaat 2406
|||||

|||||
Db 509 GGTGGGTGGCATTCTGTGCTCTGGCTTGCGCTGAACAGTTCTGGAATTACACAGCAAT 450
|||||

QY 2407 ccttgacctccctggccctgtataaattgacagcttaggactttaaacttagatcaa 2466
|||||
Db 449 CCTTGACTCCCTGCCCTGTGTAAATTGACACAGCTTAGGACTTTAAACTTAGATCAA 390
|||||

QY 2467 aagatagtcctttttaacttattttaaggagcagactttaaatgagccctgacct 2526
|||||
Db 389 AAGATATGTCCTTTTAACTTTATTTTAAGAGACGACTTTAAATGAGCCCTGACCT 330
|||||

QY 2527 ttaccattataacagaattgtcaaaaaggagtglttcttgaggaggtagcttttlla 2586
|||||
Db 329 TTACCCATTATACAGAAATTGTCAAAAGAGTGTCTTGAGGAGGTAGCTTTTTTTA 270
|||||

QY 2587 ccacactacaggacattaccctgtaggccagaagactacagctggtgtcccttagaggc 2646
|||||
Db 269 CCACACTACAGGACATTACCTGTAGGCCCAGAAAGACTACAGGCTGTGTCCTTAGAGGGC 210
|||||

QY 2647 ccaatacgtcaatccaacctctaagtcggygaagtgacaggttccgtgtgctgt 2706
|||||
Db 209 CCAATACAGTCAATTCCAACCTCTAAGTCGGGGAAGGTGACAGGTTTCCGTGCTGGT 150
|||||

QY 2707 gtgcacaagggcaggcaggtgaactgtgcctggygaagacatgtgtcctcctagtcagc 2766
|||||
Db 149 GTGCACAGGGGCGAGGCGAGGTCACTGGCCCTGGGGAAGACATTGTGGCTCCTTAGTCAGC 90
|||||

QY 2767 cctgtctccactctgtgttagctggaaccttcccactcatggaataaagaaactcac 2826
|||||
Db 89 CCTGCTTCCACTCTTGTGTAGCTGGAAACCTTCCACTCATGGAATATAGTAACCTCAC 30
|||||

QY 2827 ttctttgtcaccataaatgttaact 2855
|||||
Db 29 TTTCTTGTCCACCAATAATGTGTAATACT 1
|||||

RESULT 2
AW320695 633 bp mRNA EST 25-JAN-2000
LOCUS uc022b10.y1 NCI-CGAP Mam6 Mus musculus cDNA clone IMAGE:2609851 5'
DEFINITION similar to TR:095793 095793 STAUFPEN PROTEIN. ; mRNA sequence.
ACCESSION AW320695
VERSION AW320695.1 GI:6750239
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 633)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

FEATURES
source
1. 633
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:2609851"

```

/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

```

Query Match	20.98;	Score 596.6;	DB 89;	Length 633;
Best Local Similarity	97.78;	Pred. No. 1.2e-154;		
Matches 605; Conservative	0;	Mismatches 14;	Indels 0;	Gaps 0;

OY	1123	cccagcgccagcctgtccaaggccagcactcaaatcaagaagaagactccagtataagaaac	1182
Db	14	ccccaggccgacccctgccaaGCCCAGCCACTCAATTCAGAAGAAGACTCCAGTAAGAANAAC	73
OY	1183	caggagacgcgaagaaaagtaacgltttttagaacctagccctggggatgtaanaatygaacta	1242
Db	74	CAGGAGACGGAAAGGAAGTAACTGTTTTTGAACCTTAGCCCCGGGATGAAAATGGAACTA	133
OY	1243	gtaacaagaagcagagagtttcaggatgtcctlatcttagccatcagcagctgcaccgtygaa	1302
Db	134	GTAACAAGGACGAGAGATTTCAGGATGCCCTTATCTTAGCCATCAGCAGCTGCACGCTGGAA	193
OY	1303	tctcccatgtgtgccggaaglttgcccaagctgtcggggttagtcaagagaccaacca	1362
Db	194	TTCTCCCCTGTGCGGAAAGTTGCCAGGCTGTGGGGGTTAGTCAAGGACACCACACCA	253
OY	1363	aagatttcaccagagcgacctccaatctctgcgaagcaaacyglaactgcctatagccc	1422
Db	254	AAGATTTCACCAGGGGCGACTCCAATCTTGCCCAAGGCCAACGGTAACTGCCATGATAGCCC	313
OY	1423	gagagltgtgttacgggggcaacctcgccacagccgagacocatlttaaagagtaacatct	1482
Db	314	GAGAGTTGTTGTACGGGGCACCTCGCCACACAGCCGAGACCATTTTAAAGAGTAACATCT	373
OY	1483	cttcagggcacgtaccoccatygaacctcgcactagaccctctyagcaactgtlactacctt	1542
Db	374	CTTCAGGCCACGTACCCCGATGGAGACCTCGCAGCTAGACCTCTGAGCAACTGTACTACCTTT	433
OY	1543	ccagagcccaaggattccaggttgaatatacaagaatttcccaagaacaacaagaacqagt	1602
Db	434	CCAGAGCCCCAGGGATTCAGAGTTGAATACAAAGATGTTGCCAAGAGCAACGAGACGAGT	493
OY	1603	gtgtatctctcatcaactgtctcctcacagccgcctctcgtcagtcatgtgcactcggcaag	1662
Db	494	GTTGATCTCTCATCACTACTGCTCCTCACAGCCGGCTCTCGTCAGTCATGGCATCGGCAAGG	553
OY	1663	atgttgagtcctgtcatyatatatgtgtgcactygaacatltaaagctgtgtctgaglttg	1722
Db	554	ATGTGAAGTCTGTCTATGATATGGCTGCACCTGGACATTGTAAGCTGCTGTGTGAGTTGG	613
OY	1723	accaacagagcacagagat	1741
Db	614	ACTCACAGAACACAGATAT	632

RESULT	3				
BE867390					
LOCUS					
DEFINITION	BE867390	1018 bp	mrna	EST	27-SEP-2000
	601442470F1	NIH_MGC_65	Homo sapiens	cdna clone	IMAGE:3846689 5' ,
ACCESSION	BE867390				
VERSION	BE867390.1	GI:10316166			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1018)	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.			

Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM546 row: 1 column: 18
High quality sequence stop: 654.

FEATURES	Location/Qualifiers
source	1. .1018

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846689"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

```

Query Match	20.4%;	Score 582.6;	DB 136;	Length 1018;
Best Local Similarity	81.8%;	Pred. No. 1.1e-150;		
Matches 735; Conservative	0;	Mismatches 154;	Indels 10;	Gaps 5;

OY	925	ttgcacagatccagcagcaaaaaaggagaagagccagagtacatgcttccttacagaac	984
Db	1	TGGCCAGATCCAGCAGGCCAAAAAGGAGAAGAGGCCAGACTACACGCCTCTCACAGACC	60
OY	985	gaagcttccacactcgacggagtttgtatgcagtgtaagglttggcatcacacatgcag	1044
Db	61	GAGCGCTCCCCGCCGCCGAGGAGTTTGATGCAGGTGAAGTTGGAAAACCACTGCAG	120
OY	1045	aagagtgtygtaccataaagaagtgygccaaagcylaatyctgtgagaacatgctgaga	1104
Db	121	AAGGAACGGGGCACCAACAAGAAGTGGCCAACGCCAATTGCACGCCGAGAACATGCTGAGA	180
OY	1105	tcctygygtlcaaagtlccccagycgcagcctgccc aaagccagcactcaatcagaagaga	1164
Db	181	TCCTTGTTCAAAGTCCCAGGCGCAGCCACCCAACCGCACTCAAGTCAGAGAGA	240
OY	1165	agactccagtaaagaacccagagacggaagaaagtaacgtttttlgaacctagccctg	1224
Db	241	AGACACCCCATTAAGAACCAGGGGATGGAAGAAAAGTAACTTTTTTGAACCTGGCTCTG	300
OY	1225	gggatgaaaatggaactagtaacaagacgagaggtlccagatgacctatcctaagccatc	1284
Db	301	GGGATGAAATGGGACTAGTAATAAGAGGATGAGTTCAGGATGCCCTATCTAAGTCATC	360
OY	1285	agcagctgccagctgnaattctccocatggtgcggaagtgtgccagagctgtcgsgtta	1344
Db	361	AGCAGCTGCCCTGCTGGAAFTCTTCCCATGGTGCCCGAGGTGCCCCAGGCTGTAGAGTTA	420
OY	1345	gtcaagzacacacacccaagaattlcaccagggcagctlccaatlccctgccaagzcaag	1404
Db	421	GTC AAGGACATCACACCAAGAATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGG	480
OY	1405	taactgccatgatagcccgagagttgtgtacgysggycacctcgcccaacagccgaacca	1464
Db	481	TAACTGCCATGATAGCCCGAGAGTTGTGTATATGGGGCACCTGCCCCACAGCCGACCA	540

IMAGE Consortium (info@image.jnl.gov) for further information.
MGI:1079441
High quality sequence stop: 445.
Location/Qualifiers
1. 624

FEATURES
source
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3369837"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker: 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 153 a 138 c 155 g 178 t
ORIGIN

Query Match 19.4%; Score 554.4; DB 110; Length 624;
Best Local Similarity 96.2%; Pred. No. 6.2e-143;
Matches 600; Conservative 0; Mismatches 21; Indels 3; Gaps 3;

OY 1653 atcggaagatgtgagtcctgcatgatagtgctgcaactttaagctgtg 1712
||||| ||||||| ||| ||| ||||||||| ||||||||| |||||||
Db 624 ATCGGACGCGATGTGAGCTCTGTATGATATGCGACACTGACATTTT 565
OY 1713 tctgagttgaccaagagcacagagatgccaaagaagaatgaccagttcagcg 1772
||||| ||||| ||||| ||||||| ||||||| ||||||| |||||||
Db 564 TATGAGTGGCAGCTTACAGAGCTCAGAGATGCCAGAAACAGAGATGGCC 505
OY 1773 tgcggaggtgctgaacctttctgcccacaaccattataaaccacatat 1832
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 504 TCGGCGAGGTGCTGAACCTTTCTGGCCGCCAACCAATTATAAACGCA 445
OY 1833 aaatactggaactgcttgaataattggaatatctgataactccagtg 1892
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 444 AAAATACTGAGAACTGTTGAAATTTGGAAATATCTGATACTCCAGT 385
OY 1893 tggtagataaaatgtggcaaaagacagacaagaacttcagtgtagcc 1952
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 384 TGGTGGATAAAAATGTGCAAAAGACGACAGAACTTCAGGTGTGCTG 325
OY 1953 -ggcgctagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2011
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 324 GGGCGGCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 265
OY 2012 ctcagttctgttttgcacgtgacaaagagacagccaatttcattgt 2071
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 264 CTCCAGTCTGTTTTCATCTGTCAGCAAGACAGCAATTCATGCTGCT 205
OY 2072 tcagatacttgaanaaccgagacagccacacagagagagcctatag 2131
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 204 TCAGATACTTTGAAAAACCCGGACAGCCACAGAGAGGCTTATAGCG 145
OY 2132 taaacgagcagagaagaagcagtgcttcctactgacatgactgact 2191
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 144 TAAACGACACAGAGAAA-ACCAGTGTCTTCTACTGCACATGACTGAC 86
OY 2192 cactgtagcaccactgtaaccactgtcttctctc-agttcattttt 2250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 85 CACGTAGCACCACCTGTACCACTGCTTCTTCAAGTTTCAATTTTCT 26
OY 2251 tacaacactataatttcatttca 2274
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 25 TACAACACTATATATTTTCATTTC 2

RESULT 6
AW555746/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@jgsum.gsc.nia.nih.gov
Plate: L0259 row: A column: 06
Seq primer: -21M13 Forward
High quality sequence stop: 677.
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1. 677
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L0259A06"
/clone_lib="Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an oligo(dT) primer [NotI
primer-adaptor from GibcoBRL]
[5'-pgactgactgtctagatcgccgagcgccgctttttttttttt-3'] from
2.5ug of total RNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker LL-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao."

BASE COUNT 162 a 158 c 163 g 194 t
ORIGIN

Query Match 18.5%; Score 529.8; DB 92; Length 677;
Best Local Similarity 92.6%; Pred. No. 4.5e-136;
Matches 629; Conservative 0; Mismatches 2; Indels 48; Gaps 5;

OY 1641 gtcagtcatgcatgcgcgaagatgtgagtcctgtcatgatagtcgtc 1700
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 677 GTCAGTCATGGCAGCGCAAGAGATGTGAGTCTGTGCATGATATGGCT 618
OY 1701 ttaagctgtctgagttgaccacaagagcacagagatgccaagaacga 1760


```
|||||
Db 617 TTAAGCTGCTGTGAGTTGAGCCACAGACAGAGATGCCAAGAACAGGAATGGA 558
QY 1761 ccagttcagc-----gtgcg 1776
Db 557 CCAGTTTCAGCGGTGAGTGCACACATGCAGAGCCAATGGCAGTGTGGGTGAGCAGGTGCG 498
QY 1777 ggaagtgtgaaccttttctgtgccacaacacattataaaccacatatatactgaaaa 1836
Db 497 GGAGGTGCTGAACCTTTTCTGTGCCACACAACCATTTATAAACCCACACATATATACTGAAAA 438
QY 1837 tactgagaactgtcttgaataatttgaatatctgataactccagtgaggccaagacatgt 1896
Db 437 TACTGAGAAGCTGCTTGAATAATTGGAAATATCTGATACTCCAGTGGGCCCAAGACATGCT 378
QY 1897 gataaaaaatgtgcaaaagacgacaagaacttcaggtgtgtagccctgtgtgt-ggc 1955
Db 377 GGATAAAAATGTGGCAAGACGACAGAAACTTCAGGTGTAGCCCTGGTGTGCTGGGC 318
QY 1956 ggctagtgatgtgtgtgtctgtccatccatccagacagaacacagccccaacgctcc 2015
Db 317 GGCTAGTGATGATGCTGTGCTCTGCCATCCATCCAGACAGAAACCCCAACGCCCTCC 258
QY 2016 agtctgttttgcacgtgcacaagagagacagccaattctcatgtgtgtcttcag 2075
Db 257 AGTCTGTTTTTGCATGCTGACAAAGAGAGACAGCCAATTCTCATGCTGTGCTTCTTCCAG 198
QY 2076 atacttgaaaaaccggaacagaccacacagagagccttatagcggcccgagctaaa 2135
Db 197 ATACTTTGAAAAACCGGAGACAGCCACACAGAGAGGCCCTTATAGCGGCCCGAGCTAAA 138
QY 2136 cgaaccagagaaaagcagtgcttctcactgcacatgactgaactcagctccgacacag 2195
Db 137 CGCACCCAGAGAAAA-ACCAGTGTCTTCTTCTGACATGACTGACTGAGTCCAG-TCCGCCACACG 80
QY 2196 tagcaccactgtaacactgtcttctcttc-agtttcatlttttctcttgattgataca 2254
Db 79 TAGCACCACTGTAAACCACTGCTTCTCTCAAGTTTCATTTTTCCTTTGATTGATACATA 20
QY 2255 acactataattttcatctc 2273
Db 19 ACACTATAATTTCATTTC 1

RESULT 7
AM537709/c 686 bp mRNA EST 31-AUG-2000
LOCUS G0120F09-3 Mouse E7.5 Embryonic Portion cDNA Library Mus musculus
DEFINITION CDNA clone G0120F09 3', mRNA sequence.
ACCESSION AM537709
VERSION AM537709.1 GI:7180222
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 686)
REFERENCE Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
AUTHORS ,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and
JOURNAL embryo using a 15,000 mouse developmental cDNA microarray
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: G0120 row: F column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 686
POLYA=Yes.
```

```
FEATURES
source
Location/Qualifiers
1. 686
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G0120F09"
/clone_1ib="Mouse E7.5 Embryonic Portion cDNA Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
6 Embryo. The double-stranded cDNA was synthesized by
Gibco's kit with an Oligo(dt) primer [NotI primer-adapter
from GibcoBRL]
[5'-pGACTAGTCTAGATCGGAGCGGCCGCTTTTCTTTTCTTT-3']
from 0.5ug of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker Ll-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Minoru S. H. Ko."
BASE COUNT 214 a 139 c 150 g 183 t
ORIGIN
Query Match 18.4%; Score 526.6; DB 91; Length 686;
Best Local Similarity 99.1%; Pred. No. 3.5e-135;
Matches 540; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 2313 ctgtcagtcattgaactggaagatcacagccgggttggtgcatctgtgcctctg 2372
Db 686 CTGTCACTCATGTACTGGCAAGGATCACAGCCGGTGGGTGGCAATTCTGTGCTCGGG 627
QY 2373 ctggctgaacagttcttgaattaccaccagaatccttgaactccctgcccctgtataaa 2432
Db 626 CTTGGCTGAACAGTTCTGGAATTACCAACAGAAATCCTTGACTCCCTGCCCTTGATATAA 567
QY 2433 ttggacagcttagacctttaaactttagatcaaaagatatgtctctttaaacttaatt 2492
Db 566 TTGGACAGCTTAGACTTTTAACTTTAGATCAAAAAGATATGTCCTTTTAACTTATAT 507
QY 2493 tttaagagcagactttaaatagccctgaccttaaccattataacagaatttgtcaa 2552
Db 506 TTTAAGGAGCAGACTTTTAAATGAGCCCTGACCTTTTACCATTATATACAGAAATTGTCAA 447
QY 2553 aaggagtgcttcttgaagagtgagcttttttaccacactacagagacattacctgtag 2612
Db 446 AAGGAGTGTTTCTTGAGGAGGTAGC-TTTTTTTACCACACTACAGAGACATTACCTGTAGG 388
QY 2613 cccagaagactacagcgtgtgtccctagagggcccaatacagtgcaattccaacctcaa 2672
Db 387 CCCAGAAGACTACAGGCTGTGTCCCTTAGAGGGCCCAATACAGTCAATTCCAACCTTAA 328
QY 2673 gtcggggaaggtgacaggttctcgtgtgtgtgtgtgcacaggggacaggtcagctg 2732
Db 327 GTCGGGGAAGAGGTGACAGGTTTCTGTGTGTGTGTGCACAGGGGCAAGTCAGCTG 268
QY 2733 gctctggggaagagcatgtgtgtcctcctagtgcagccctgtctccactctgttagctg 2792
Db 267 GCCTGGGGAAGAGCATGTGGCTCCTAGTGCAGCCCTGCTTCCACTCTTGGTTAGCTGG 208
QY 2793 aacctccactcatggaataataagtaactcaacttcttgtcacaataaatgtgtaac 2852
|||||
```

Dn	207	AACCTTCCCACTCATGTGAAATATAAGTAACACTCATTTCCTTTGTCCACCATAAATGGTAAT	148
OY	2853	actaa 2857 	
Dn	147	ACTAA 143	
RESULT	8		
LOCUS	AAI20063	564 bp mRNA	EST 17-FEB-1997
DEFINITION	mp88b08.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:576279		
ACCESSION	AAI20063		
VERSION	AAI20063.1	G1:1677763	
KEYWORDS	EST.		
SOURCE	house mouse. Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 564)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:350927		
FEATURES	Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 488.		
SOURCE	Location/Qualifiers 1..564 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:576279" /clone_lib="Soares_thymus_2NbMT" /sex="male" /tissue_type="Thymus" /dev_stage="4 weeks" /lab_host="DH10B" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGGCCGCTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	150 a 142 c 124 g 148 t		
ORIGIN			
Query Match	18.4%; Score 525.8; DB 2; Length 564;		
Best Local Similarity	99.1%; Pred. No. 5.4e-135;		
Matches 560; Conservative 0; Mismatches 2; Indels 3; Gaps 3;			
OY	1831	tgaatactagagaactgcttggaaaatttggaatatctgataactccagtggccaaga	1890
Dn	1	TGAAAATACTGAGAAGCTGCTTTGAAAAATTGGAAATATCTGATTAACCTCCAGTGGCCAAGA	60
OY	1891	catgtgtgataaaaaatgttcgcaaaagacgcacaagaacctcaggtgtgtagccctgtgtgtg	1950

Db	61	CATGGTGGATAAAAAATGTGGCAAGACGACAGAAACCTTCAGGTGGTAGCCCTGGTGTG	120
QY	1951	ct-ggcgcctagtgatgatgtgtgtctctgcatccatccagacagaaccagccccaac	2009
Db	121	CTGGGGCGGCTAGTGATGATGCTGTGCTCTGCCATCCATCCAGACAGAAACCAAGCCCAAC	180
QY	2010	gcctccagttctgttttttgatcgtgtgacaaagagagacagccaattctcattgtgctt	2069
Db	181	GCCTCCAGTCTGTGTTTTGCATCGTGACAAGAAGAGAGACAGCCCAATTCTCATGCTGGCTT	240
QY	2070	cttcagatactttgaaaaaaccccgacagccacaccagagagccttatagcggcccgga	2129
Db	241	CTTCAGATACTTTGAAAAAACCCGGACAGCCACACCAAGAGAGGCTTATAGCGGCCCGGA	300
QY	2130	gctaacaggaaccagaaaagccagtgcttcctactgacatgactgactcagctccgc	2189
Db	301	GCTAAACGCACACAGAGAAAAA-CCAGTGCCTTCTACTGCACATGACTGACTCAGCTCCGC	359
QY	2190	cacacgtagaccaccactgttaaccactgtcttcctc-aggttcatlttttcccttgatt	2248
Db	360	CACACGTAGCACCACTGTAAACCACTGCTTCTCTTCAAGTTTCATTTTTCCTTGTGATT	419
QY	2249	gatacaacactataatttcatcttcagttccttagtcgtgtctacttacctagcagttta	2308
Db	420	GATACAACACTATAATTTTCATTTCACTTCAGTTCCTTAGTCGTGCTACTTACCTAGAGTTTA	479
QY	2309	gaaactgtcagtcattgttaactggaagatcacagcccggttggtgycattctgtgcct	2368
Db	480	GAAACTGTCTAGTCATGTAACTGGCAAGATCACAGCCCGGTGGGTGGCATTCGTGGCT	539
QY	2369	ctggtctggtgaacagttctgga	2393
Db	540	CTGGCTTGGCTGAACAGTCTTGGA	564

```

RESULT      9
LOCUS       BF016149
DEFINITION  BF016149      522 bp      mRNA
            uy31e08.y1 NCI-CGAP_Lu30 Mus musculus cDNA clone IMAGE:3659630 5'
            similar to TR:095793 095793 DOUBLE-STRAND RNA-BINDING PROTEIN
            STAUEN. ; , mRNA sequence.
ACCESSION   BF016149
VERSION     BF016149.1  GI:10747481
KEYWORDS    EST.
SOURCE      house mouse.
            Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 522)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

MGI:1421958
Seq primer: -40RP from Gibco
High quality sequence stop: 441.
Location/Qualifiers
1..522
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"

```

```

/clone="IMAGE:3659630"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; transgenic model WNT-1, expression driven by
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"

```

Query Match	18.1%;	Score 517.2;	DB 139;	Length 522;
Best Local Similarity	99.4%;	Pred. No. 1.3e-132;		
Matches 519; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

OY	81	gcgtctgttgtagcgtccccccttcctctgagccccggcctgycgycgccgccttcgcctc	140
Db	1	GGCTGTGTGTCGCCTCCCCCCTTCTTGAGCCCCCGCCTGGCGGCCGCTTTCGCCTC	60
OY	141	cggcaactcgcgectcttcacctctctggtcgcccttttccctgcgctctcaattgct	200
Db	61	CGCCACTCCGCTCTTCCCTCTCCGGTCGGTCCCTTTTTCCTCGCCGCTTCTTGCT	120
OY	201	tcttcacctctctcgccgcccgaagaccgcccggaggacgactctgggaaagcag	260
Db	121	TCTTCAACCTCTCGCCGCCGCCAAGACC GCCGCCGCCGACGAGCTCTGGGGAAGCAG	180
OY	261	ccagaagaatagcttctaaccattgaagctcaatgcactgtgtgaaactgaaaagaaaa	320
Db	181	CCAGAAAGATATAGCTTCTACCATTTGAGCTCAATGCACGTGTGTGAACCTGGAAGA	240
OY	321	ccaatgtataagcccgtygaccctcactctcgatgcagtcaccactacagtatgcaty	380
Db	241	CCAATGTATAGCCCGTGAGCCCTCACTCTCGGATGCAGTCCACCTACAGCTATGSCATG	300
OY	381	cgtgagagtgcctataccccccagataactttaaccattcccaagtcaccacttaacttac	440
Db	301	CGTGAGGTGCTTATCCCCCCAGATATTTTAACCATTTTCCAGTCCCACCTTTACTCTAC	360
OY	441	caagltgagctctccgtlygcygacagcagltlaatygaagaaagaaagatgagacaccc	500
Db	361	CAAGTTGAGCTCTCCGTGGGGGACAGCAGTTTAATGGGAAAGAAAGATGAGACACCC	420
OY	501	gtgaacaacgagtcgccctcgcccgtgcgctgagagactctgcagagtgaaacccctgcagaa	560
Db	421	GTGAACAACGATGCCCTTGCCCGTGCCTGAAGACTCTGCAGAGTGAACCCCTGCCAGAA	480
OY	561	aggttgaggttaaatggaagagaagcagaaggaagaaaccttc	602
Db	481	AGGTGAGGTTAATGGAAGAGAAGCAGAGGAAGAAAACCTTC	522

RESULT	10
AAI122533	
LOCUS	
DEFINITION	AAI122533 522 bp mRNA EST 17-FEB-1997 mp95c04.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:576966 5' , mRNA sequence.
ACCESSION	AAI122533
VERSION	AAI122533.1 GI:1681596
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 522)
REFERENCE	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R.
AUTHORS	The WashU-HHMI Mouse EST Project
TITLE	

JOURNAL COMMENT	Unpublished (1996) Contact: Maria M/Mouse EST Project
-----------------	--

Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:351614
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 448.

FEATURES	Location/Qualifiers
source	1. .522

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:576966"
/clone_1lb="soares_thymus_2nbmr"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Ronaldo."

```

Query Match	17.9%;	Score 511;	DB 2;	Length 522;
Best Local Similarity	99.8%;	Pred. No. 6.9e-131;		
Matches 522; Conservative	0;	Mismatches 0;	Indels 1;	Gaps 1;

QY	2248	tgaatacaacaactataaatttcattcattcagttcccttagtcgtctacttaacctagcaattt	2307
Db	1	tgatatacaacactataatttttcatttcagttcccttagtcgtctacttaacctagcaattt	60
QY	2308	agaacatgtcagtcacatgttaactgtgcaagatcacagaccggttgglygcattctgtgcc	2367
Db	61	agaacatgtcagtcacatgttaactgtgcaagatcacagaccggttggggcattctgtgcc	120
QY	2368	tctggtctgctgaacagttcttgaattaccaccagaatcccttgaactccctgcccctgt	2427
Db	121	tctgtg- ttggctgaacagttcttgaattaccaccagaatcccttgaactccctgcccctgt	179
QY	2428	ataaattgacacagcttaggacttttaaccttagatcacaagaatatgttccttlttaact	2487
Db	180	ataaattgacacagcttaggacttttaaccttagatcacaagaatatgttccttlttaact	239
QY	2488	ttaattttaaggagcagactttaaaatgagccctgaccttaccattatacagaattt	2547
Db	240	ttatattttaaggagcagactttaaaatgagccctgaccttaccattatacagaattt	299
QY	2548	gtcaaaaaggagtgttctcttgaggaggtagccttttltaccacaactacagacataacct	2607
Db	300	gtcaaaaaggagtggttcttgaggaggtagccttttlttaccacaactacagacattacct	359
QY	2608	gttagccccagaagaactacagagctgtgtccctagaagggcccaatacagtcattccaacc	2667
Db	360	gttagccccagaagaactacagagctgtgtccctagaagggcccaatacagtcattccaacc	419
QY	2668	tctaagtcggggaaggtgacaggttctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2727
Db	420	tcttagtcggggaaaggtgacaggttctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	479

```
QY 2728 agctggcctggggaagagcattgtgtcctcctagtcagccctg 2770
|||||
Db 480 AGCTGGCCCTGGGGAAGAGCATGTGCTCCTAGTGCAGCCCTG 522

RESULT 11
LOCUS BE894295 752 bp mRNA EST 29-SEP-2000
DEFINITION 60143766ff1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922529 5',
mRNA sequence.
ACCESSION BE894295
VERSION BE894295.1 GI:10356518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM744 row: d column: 18
High quality sequence stop: 654.

FEATURES
source location/Qualifiers
1..752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922529"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 223 a 194 c 192 g 143 t
ORIGIN
Query Match 17.8%; Score 508.4; DB 137; Length 752;
Best Local Similarity 84.8%; Pred. No. 4.1e-130;
Matches 616; Conservative 0; Mismatches 106; Indels 4; Gaps 4;

QY 892 attatgccaaggatgaatcctatagtagactgcacagatccagcgcaaaaagg 951
|||||
Db 1 AATATGGCCAGGGGATCAATCCGATTAGCCGACTGGCCAGATCCAGGCAAAAAAGG 60

QY 952 agaagagccagagtagatgtcctctacagaacgaggtcttccacgtcgcagaggtttg 1011
|||||
Db 61 AGAAGGAGCCAGAGTACACGCTCCTCACAGAGCGAGGCTCCCGCGCGAGGAGTTTG 120

QY 1012 tgatgcaggtaaaggttggtgcatcacactgcagaagagtggtaccataagaagttg 1071
|||||
Db 121 TGATGCAGGTGAAGTTGGAACCACTGCAGAGAAGGAGCGGACCAACAAGAAGTTGG 180

QY 1072 ccaagcgtaatgtctgtgagaacatgtgtgagatcctgtggttcaagtctcccgagcgc 1131
|||||
Db 181 CCAAGCGCAATGCAGCCGAGACATGCTGAGATCCTTGTTTCAAGTCCCGCAGGCGC 240

QY 1132 agcctgcgaagccagcactcaaatcaagaagaagactccagttaagaaccagagacg 1191
|||||
Db 241 AGCCACCAAAACCCGCACTCAAGTCAAGAGGAGAGAACACCACTAAAGAAACCAAGG-GATG 299
```

```
QY 1192 gaagaaagtaacglttttttgaacctagccctgggagatgaanaatggaactagtaacaag 1251
|||||
Db 300 GAAGAAAGTAACCTTTTGTGAACCTGCTCTGGGGATGAATAATGGACTAGTAATAAG 359

QY 1252 acgagagttcaagatgtcctatctatcctacatcagcagctccagctgtgaattctccca 1311
|||||
Db 360 AGGATGAGTTCAGGATGCGCTTATCTAAGTCATCAGCAGCTGCTGTGAATTCTTCCCA 419

QY 1312 tgtgccccgaagttgccagagctgtcgggttagtcaagagacacacacaaagattca 1371
|||||
Db 420 TGTGCCCGAGGTGCGCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGATTTTA 479

QY 1372 ccaaggcagctccaatcctctgccaaggaacggttaactgcatgatagcccagagttgt 1431
|||||
Db 480 CCAGACAGCTCCGAATCCTGCCCCAAGGCCACGGTAACCTGCCATGATAGCCCCGAGATTGT 539

QY 1432 tgtacgggggacactcgccacagccgagaccattttaa-gagtaacatctctcagc 1490
|||||
Db 540 AGTATGGGGGCACCTCGCCACAGCCGAGATCATTTTAAACGATAACATCTCTCAGGC 599

QY 1491 caggtaccctatgacctgcacatagacctctgagcaactgtactaccttccagagcc 1550
|||||
Db 600 CACGTACCCCATGGA-CTCTCACGAGATCTCTGAGCAA-TGGACTATCTTCCAGAGTC 657

QY 1551 caggatccaagttgataacaagatttccaagaacaagaacgaagtgatctatct 1610
|||||
Db 658 CAGGATTCGGGTGTATACAAGACTTTCCAAAAAACAAGAGAATTTGTTCTCTATCA 717

QY 1611 ctcatc 1616
||
Db 718 ATGCTC 723
```

```
RESULT 12
AW542549/c 511 bp mRNA EST 31-AUG-2000
LOCUS AW542549/c
DEFINITION C0158H10-3 Mouse E7.5 Extraembryonic Portion cDNA Library Mus
musculus cDNA clone C0158H10 3', mRNA sequence.
ACCESSION AW542549
VERSION AW542549.1 GI:7184966
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 511)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Dol,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE 20381348
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: C0158 row: H column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 511
POLYA=Yes.

FEATURES
source location/Qualifiers
1..511
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C0158H10"
/clone_lib="Mouse E7.5 Extraembryonic Portion cDNA
library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
```

/lab_host="DH10B"
/note="vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
5 EPC. The double-stranded cDNA was synthesized by
Gibco's kit with an Oligo(dT) primer (NotI primer-adapter
from GibcoBRL)
[5'-pGACTAGTTCATGATCGCAGCGCGCCCTTTTCTTTT-3']
from 0.8ug of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker L1-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang."

BASE COUNT 149 a 120 c 116 g 126 t
ORIGIN

Query Match 17.8%; Score 507.8; DB 92; Length 511;
Best Local Similarity 99.6%; Pred. No. 5.3e-130;
Matches 509; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2345 ccggttggtgcatctctgtcctcgtcgtcgtgaacagttctggaataccaccaga 2404
Db 511 CCGGTTGGGTGGCATTCGTGCTCTGCGTTGGCTGAACAGTTCGGAATTAACACAGCA 452
OY 2405 atccttgactcgcctgccccctgtataaaltgagacagcttagactttaacttagatc 2464
Db 451 ATCCTTGACTCCCTGCCCCCTGTATAAATTGGACAGCTTAGGACTTTTAACTTAGATC 392
OY 2465 aaaaatatgtctcttttaacttattttaagagagcagacttaaatgagccctgac 2524
Db 391 AAAAGATATGCTCCTTTTAACTTTATTTAAGAGACAGACTTTAAATGAGCCCTGAC 332
OY 2525 cttaaccattataacagaattgtcaaaagagtgcttcttgagaggtagctttttt 2584
Db 331 CTTTACCCATTATACAGAAATTGTCAAAAGAGTGTCTTGAGAGAGTCTTTT 272
OY 2585 taccacactacagacattcctgtagggccagaaactacagagctgtgtccctagag 2644
Db 271 TACCACACTACAGGGCAATTAAGTGTAGGCCACAGAGACTACAGGCTGTGCTCTAGAGG 212
OY 2645 gcccaatacagtcacactcctaagtcggggaaggtgacaggttccctgtgctg 2704
Db 211 GCCCAATACAGTCAATTCACACTTAAGTCGGGAAAGGTGGCAGGTTCCGTGCTG 152
OY 2705 gtgtgcacaggggcaggtcagctgagctggggaagagcatgtgtgctcctagtga 2764
Db 151 GTGTGCACAGGGGCAGGCACTGAGCTGGGGAAGAGCATTTGTGCTCTAGTGCA 92
OY 2765 gccctgctccactctgtgttagctggaacctcccaactcatggaataagtaactc 2824
Db 91 GCCCTGCTTCCACTCTGTGTTAGCTGGAACCTTCCCACTCATGGAATATAAGTAAACTC 32
OY 2825 accttcttgtcaccataaatgttaatact 2855
Db 31 ACTTCTTTGTCCACCAATAATGTATAACT 1

RESULT 13
LOCUS BE674157 616 bp mRNA EST 08-SEP-2000
DEFINITION 7d76a04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3278862 3'
ACCESSION BE674157 similar to TR:095793 095793 STAUZEN PROTEIN.; mRNA sequence.

VERSION BE674157.1 GI:10034698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 616)
JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 469.
Location/Qualifiers

FEATURES
source 1. 616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3278862"
/tissue_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker. Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 179 a 153 c 139 g 141 t 4 others
ORIGIN

Query Match 17.6%; Score 502.2; DB 111; Length 616;
Best Local Similarity 89.0%; Pred. No. 2e-128;
Matches 540; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 1164 aagactccagtaaagaacaccagagacggaagaaagtaacgtttttgaacctagccct 1223
Db 8 AAGACACCCATAAAGAAACACAGGGGATGGAAGAAAGTAACCTTTTGAACCTGCTCT 67
OY 1224 ggggatgaaatggaactagtaacaagagcagaggttcaggtatgcctatcttagccat 1283
Db 68 GGGGATGAAATGGGACTAGTATTAAGAAGAGATGAGTTCAGCATGCCCTATCTAGTCAT 127
OY 1284 cagcagctgccagctggaatctcccatggtgcccgaaggtgcccaggtgctggggtt 1343
Db 128 CAGCAGCTGCTGCTGGAATTTTCCCATGTGTGCCGAGGTGCCCCAGGCTGTAGAGTT 187
OY 1344 agtcaagacacacacacaaagaattcaccaaggcagctccaatctctgccaagcaag 1403
Db 188 AGTCAAGGACATCACACCAAGATTTTACAGGGCAGCTCCGAATCTGCAAGGCCACG 247
OY 1404 gtaactgcataatagcccgagagtggtgtacaggggacactcgccacagccgagacc 1463
Db 248 GTAACGTCCATGATAGCCCGAGAGTTGTGTATGGGGCACCCTCGCCACAGCCGAGACC 307
OY 1464 atttaagagtaacatctcttcagggcagcgtaccccatgacactcgacactagacctct 1523
Db 308 ATTTTAAAGATAACATCTCTTCAAGGCCACGTAACCCCATGAGACCTCTCACAGAACCTCT 367

QY	1524	gagcaactgtactaccttccagagcccaaggatccaggltgaatacaagaatttccc	1583
Db	368	GAGCAACTGTGACTATCTTTCCAGAGTCCAGGATTCAGGTTGAATACAAAGACTTCCCC	427
QY	1584	aagaacacaagaagagtggtatctctcatcaactgtctcaccagccgctctcgtc	1643
Db	428	AAAAACAACAAGAAGCAATTTGTATCTCTTATCAATTTGCTCTCANCCACCTTGATC	487
QY	1644	agtcatgcatcgcgaagatgtgagtcctgtcatgatatgctgcactgaacattta	1703
Db	488	AGCCATGGTATCAGCAAGATGTGAGTCTCCCATGATATGGCTGCGCTGAACATCTTA	547
QY	1704	aagctgtctgtcgtgagttggaaccaacagagcacaagagatgccagaacaggaatgacca	1763
Db	548	GAGTTGCTGTCTGAATGTGACCAACAANTACAGAGATGCCAAGAACANGANACAGACCA	607
QY	1764	gtttcag	1770
Db	608	ATGCTCTG	614
RESULT	14		
LOCUS	W89457	520 bp	MRNA
DEFINITION	mf73e03.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA		
ACCESSION	W89457		
VERSION	W89457.1	GI:1541890	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 520) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The Washu-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1404827. Contact: Marra M/Mouse EST Project Washu-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:254484 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 491. Location/Qualifiers 1.520 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:419932" /clone_lib="Soares mouse embryo NbME13.5 14.5" /sex="unknown" /tissue_type="embryo" /dev_stage="13.5-14.5dpc total fetus" /lab_host="DH10B" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and		
FEATURES	source		

BASE COUNT	155 a	124 c	128 g	113 t
ORIGIN	cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "			

Query Match	17.58;	Score 499.4;	DB 147;	Length 520;
Best Local Similarity	99.68;	Pred. No. 1.2e-127;		
Matches 511; Conservative	0;	Mismatches 1;	Indels 1;	Gaps 1;

QY	1638	ctcgtcagtc	atgcatg	gcgcg	gaagga	atg	tgtag	tcctc	gtcat	gatata	gtc	tgca	ctg	aaac	1697							
Db	1	CTCGTCAGT	CATGGCA	TGGCAAG	GATGTG	GAGTCT	GTTCAT	GATATG	GGCTGC	ACTGAAC					60							
QY	1698	attttaag	ctgctg	ctcga	lttg	gaacca	gagaca	gagatg	cca	agaa	cagg	aaat			1757							
Db	61	ATTTAAAG	CTGCTG	CTGTGA	GTGGAC	CAACAG	AGACAG	AGATG	CCAA	GAAC	AGGAAT				120							
QY	1758	ggaccag	lttcag	cgtgc	ggga	gltg	ctga	acctt	ctc	tgcc	aaac	cat	ataa	aac	1817							
Db	121	GGACCA	AGTTT	CAGCGT	GGGAGG	TGCTGA	ACCTTT	CTTG	GGCC	ACAAC	CAATTA	AAAAAC			180							
QY	1818	ccaacata	tatact	gaaaaa	ta	ctg	gaact	gctt	tg	aaaat	ltg	gaata	tatc	tgta	actc	1877						
Db	181	CCAACA	TATAT	ACTG	AAAAAT	ACTG	AGA	ACTG	CTTTG	AAAAAT	TGGA	TATCT	GATTA	CTC	240							
QY	1878	cagtg	ggcca	agacat	g	gtg	ataa	aaat	gt	tg	cca	aa	gac	gaca	aaactc	ag	gt	g	1937			
Db	241	CAGTGG	GGCCA	AGACAT	GGTG	GTGATTA	AAATGT	GGCA	AA	GACG	ACAAG	AAAC	TTT	CAGG	TG	300						
QY	1938	agccctg	gtgtg	ct-g	gcg	ctag	tgat	gatg	atg	tg	ctc	tg	ccat	ccac	ag	1996						
Db	301	AGCCCT	GTGTG	CTGCGG	CGGTAG	TGATG	ATGCTG	CTCTG	CCATC	ATC	AGAC	AGACA			360							
QY	1997	aaccag	cccca	gcgc	ctc	ga	ttc	gtl	ttt	gc	at	cg	tg	aca	aa	gag	ag	ag	cc	aat	2056	
Db	361	AACCAG	CCCCA	AGCCTC	CACTT	CTGTTTG	CATCGT	GCAC	AAAG	AGAG	CACAG	CACTT			420							
QY	2057	ctcatg	ctg	ctctc	tc	tcag	ata	ctt	tg	aaaa	acc	cg	ga	gc	cc	ac	ag	ag	ag	gc	cta	2116
Db	421	CTCATG	CTGCTT	CTTCAG	ATACTT	TGAAAA	ACCCG	GACG	CCAC	AC	CAG	AGAG	GCCTTA		480							
QY	2117	taccg	gcccc	gcg	ag	gcta	aa	acg	g	ac	g	ac	g	a	a	a	a	a	a	a	2149	
Db	481	TAGCGG	CCCCG	GAGCTA	AAACG	CA	CCAG	AGAAAA							513							
RESULT 15																						
AM541590/c		497 bp	mrna			EST																

Email: cdna@lgsun.grc.nia.nih.gov
Plate: C0146 row: H column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 497
POLYA=Yes.

FEATURES

source

Location/Qualifiers

1. 497

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="C0146H10"

/clone_lib="Mouse E7.5 Extraembryonic Portion cDNA library"

/sex="unknown"

/dev_stage="7.5dpc Embryo"

/lab_host="DH10B"

/note="Vector: pSPORT1 (Gibco/BRL Life Technology);
site_1: SalI; site_2: NotI; Total RNAs were extracted from
5 EPC. The double-stranded cDNA was synthesized by
Gibco's kit with an Oligo(dT) primer [NotI primer-adapter
from GibcoBRL]

[5'-pgACTAGTCTAGATCGGAGCGGCCCTTTT-3']

from 0.8ug of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker LI-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang."

BASE COUNT 145 a 111 c 113 g 128 t
ORIGIN \

Query Match 17.0%; Score 486; DB 92; Length 497;
Best Local Similarity 99.8%; Pred. No. 6.1e-124;
Matches 497; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2358 attctgtgcctctgctgctgaacagttctggaattaccaccagaatccttgactccc 2417
|||||
Db 497 ATTCTGTGCCCTCTGGCTTGACAGTCTTGGAATTACCACCAGAATCCTTGACTCCC 438

QY 2418 tgcacctgtataaattgacagcttaggactttaaactltagatcaaaagatatgctc 2477
|||||
Db 437 TGCCCTTGATTAATTTGACAGCTTAGGACTTTTAACTTTAGATCAAAAGATATGCTC 378

QY 2478 ctlttaacttattttaaaggagcagactttaaatgagccctgaccttaaccattat 2537
|||||
Db 377 CTTTAACTTTATTATTAAAGAGCAGACTTTAAATGAGCCCTGACCTTTACCATTTAT 318

QY 2538 aacagaatttgcanaaaggagtgttcttgaggagtagctttttttaccacactacag 2597
|||||
Db 317 AACAGAATTGTCAAAAGAGAGTCTTCTTGAGGAGAGTAGC-TTTTTTTACCACACTACAG 259

QY 2598 gacattacctgtagggccagaagaactacagagctggtgtcccttagagggccaataacagtc 2657
|||||
Db 258 GACATTACCTGTAGGCCCAGAGAAGACTACAGGCTGTGTCTCCTAGAGGGGCCCAATACAGTC 199

QY 2658 aattccaacctctaagtcggggaagtgacaggttctcctggtgtgtgtgcacagggg 2717
|||||
Db 198 AATTCCAACCTCTAAGTCGGGGAAGGTGACAGGTTTCTGCTGTGTGTGCACAGGGG 139

QY 2718 caggcaggtcagctggtggggaagagcatgtgtgctcctagtgacgacctgtctccac 2777
|||||
Db 138 CAGGCAGGTCAGCTGGCTGGGGGAAGAGCATGTGGCTCCTAGTGACAGCCCTGCTTCAC 79

QY 2778 tcttglttagctggaaccttcccaactcatgygaatatagtaaacctcacttcttgtca 2837
|||||
Db 78 TCTTGGTTTAGCTGGAACCTTCCCACTCATGGAATATAAGTAAACTCACCTTCTTGTCA 19

QY 2838 ccaataatggtactact 2855
|||||
Db 18 CCAATAATGCTAATACT 1

Search completed: April 5, 2001, 00:30:47
Job time: 53143 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2001, 09:45:04 ; Search time 11371 Seconds
(without alignments)
1936.289 Million cell updates/sec

Title: US-09-316-048-1
Perfect score: 3142
Sequence: 1 acttcctgccggctgcggg.....taataactaaaaaaaaa 3142

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp14:*
121: em_estro15:*
122: em_estro16:*
123: em_estro17:*
124: em_estro18:*
125: em_estro19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: gb_gss7:*
159: gb_gss8:*
160: gb_gss9:*
161: em_gss5:*
162: em_gss6:*
163: em_gss7:*
164: em_gss8:*
165: em_gss9:*
166: em_gss10:*
167: em_gss11:*
168: gb_gss10:*
169: gb_gss11:*
170: em_gss12:*
171: gb_gss12:*
172: gb_gss13:*
173: gb_gss14:*
174: gb_gss15:*
175: gb_gss16:*
176: gb_gss17:*
177: gb_gss18:*
178: gb_gss19:*
179: em_gss13:*
180: gb_gss20:*
181: gb_gss21:*
182: gb_gss22:*
183: gb_gss23:*
184: gb_gss24:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1757.8	55.9	1801	146	U69197	U69197 U69197 Soar
2	701	22.3	1018	136	BE867390	BE867390 601442470
3	646.2	20.6	1036	137	BE889880	BE889880 601512132
4	621.2	19.8	724	97	AW952360	AW952360 EST364430
5	617.2	19.6	752	137	BE894295	BE894295 601437666
6	606.4	19.3	821	107	BE379687	BE379687 601159257
7	604	19.2	609	107	BE380123	BE380123 601159257
8	603.8	19.2	686	110	BE613972	BE613972 601504093
9	602.2	19.2	674	38	AV707406	AV707406 AV707406
10	596.4	19.0	829	106	BE300155	BE300155 600944602
11	594.4	18.9	616	111	BE674157	BE674157 7d76a04.x
12	588.2	18.7	594	40	AM162002	AM162002 au72c03.x
13	571.4	18.2	612	38	AV695897	AV695897 AV695897
14	565.4	18.0	749	110	BE617223	BE617223 601441906
15	555	17.7	609	136	BE810330	BE810330 MR4-PT005
16	551.6	17.6	615	134	BE082712	BE082712 RC2-BT064
17	550	17.5	550	107	BE348298	BE348298 hw16b01.x
18	548	17.4	564	40	AM163206	AM163206 au97f05.x
19	542.4	17.3	584	136	BE813737	BE813737 RC1-BN003
20	534	17.0	673	135	BE781223	BE781223 601469159
21	531.2	16.9	536	136	BE837505	BE837505 RC2-FN009
22	531	16.9	866	136	BE868328	BE868328 601443887
23	530.2	16.9	573	3	AA206573	AA206573 zq51f08.r
24	524.8	16.7	831	106	BE299828	BE299828 600944602
25	517.8	16.5	910	138	BE966164	BE966164 601660093
26	513	16.3	543	134	BE089861	BE089861 RC5-BT070
27	511.4	16.3	513	13	AA889669	AA889669 ak54a04.s
28	510.4	16.2	512	27	AI983007	AI983007 wt46g02.x
29	508.8	16.2	852	137	BE872137	BE872137 601446238
30	500.6	15.9	673	109	BE541462	BE541462 601067913
31	497.4	15.8	633	89	AW320695	AW320695 uc22b10.y
32	488.2	15.5	569	3	AA191622	AA191622 zp81g10.s
33	487	15.5	516	25	AI819766	AI819766 wj42a10.x
34	483.4	15.4	485	25	AI802592	AI802592 w16a03.x
35	478.4	15.2	480	134	BE042598	BE042598 ho26e03.x
36	475	15.1	494	89	AW351909	AW351909 RC0-CT020
37	470.4	15.0	509	142	N31181	N31181 yx64f02.r1
38	470	15.0	517	7	AA427366	AA427366 zw33c05.r
39	470	15.0	517	92	AW580332	AW580332 IL2-HT044
40	467.4	14.9	703	137	BE888420	BE888420 601514177
41	465.6	14.8	686	40	AM160378	AM160378 au72c03.y
42	457	14.5	724	38	AV700494	AV700494 AV700494
43	455.2	14.5	561	141	H16250	H16250 ym13d10.s1
44	451.8	14.4	455	96	AW874257	AW874257 hp96g04.x
45	450.4	14.3	452	17	AI190464	AI190464 qd38b02.x

ALIGNMENTS

RESULT 1
LOCUS U69197 1801 bp mRNA
DEFINITION U69197 Soares infant brain IN1B Homo sapiens CDNA clone 22368, mRNA
ACCESSION U69197
VERSION U69197.1 GI:2739420
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1801)
AUTHORS Volorio,S., Simon,G., Repetto,M., Cucciardì,M., Banfi,S., Borsani
,G., Ballabio,A. and Zollo,M.
TITLE Sequencing analysis of forty-eight human image cDNA clones similar
to drosophila mutant protein
JOURNAL DNA Seq. 9 (5-6), 307-315 (1998)
MEDLINE 99452388
COMMENT Contact: Zollo, Massimo
Telethon Institute of Genetics and Medicine
Via Olgettina 58, Milan, MI 20132, Italy
Email: zollo@tigem.it.
FEATURES
source
1..1801
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p13.1"
/clone="22368"
/clone_lib="Soares Infant brain 1N1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site.1: Not
I; Site.2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGGAAGAAATTCGGCGCCGAGGAATTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 503 a 409 c 386 g 503 t
ORIGIN

Query Match 55.9%; Score 1757.8; DB 146; Length 1801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1371 gctcgaatcctgcgaagcgcaagtgtaactgcatgatagccgagagtggtgatgg 1430
|||||
Db 5 GCTCCGAATCCGCGCAAGGCCACGCGTAATGCCATGATAGCCCGAGAGTTGTATGGG 64
QY 1431 ggcacctgcgccacagcgagaccatttaagaataacatctctcaggccacgtacc 1490
|||||
Db 65 GGCACCTCGCCACAGCCGAGACCATTTTAAAGAAATAACATCTCTTCAGGCCACGTACC 124
QY 1491 catgacctctcaagagaccctctgagcaactgactatcttccaaggtccaaggtatc 1550
|||||
Db 125 CATGGACCTCTCAGCAGACCCCTCTGAGCACTGACTATCTTCCAGAGTCCAGGGATTTC 184
QY 1551 caggltgaatacaagacttccccaaaaacaagaacgaatttgatatcttatacaat 1610
|||||
Db 185 CAGGTTGAATACAAGACTTCCCAAAAACAACAAGAAATTTGTATCTCTTATCAAT 244
QY 1611 tgcctctcagccaactctgatacagccatggtatcggcaagatgtygaagtcctgccaat 1670
|||||
Db 245 TGCTCTCTCAGCCACTCTGTATCAGCCATGTGTATCGGCAAGGATGTGGAGTCTGCCAT 304
QY 1671 gatagtgctgcgtgaacatcttaagttgctgctgagttggaaccaacaagtacagag 1730
|||||
Db 305 GATATGGCTGCGCTGACATCTTAAAGTTGCTGTGAGTTGGACCAACAAAGTACAGAG 364
QY 1731 atgccaagaacaggaacggaacatgctgtgtgtggaagtgctgaacctttctgac 1790
|||||
Db 365 ATGCCAAGAACAAGCAACGACCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGC 424
QY 1791 catgaacctataaatacccaatatatactgaaataactgaactgctttgaaatt 1850
|||||
Db 425 CATGAACCATTTATAAATCCCAACATATATCTGAATAATACTGAACACTGCTTGAATAAT 484

QY 1851 tggaaattctgatacctccagtgccgagagagacaagtggtgtaaaagatgtggcagca 1910
|||||
Db 485 TGGAAATTTCTGATTAACCTCCAGTGGGCCGAGACACACGGTGGTTAAAGATGTGGGCAGCA 544
QY 1911 gcagggaagacaagaagaacacaagagcgctgtggcc-ggctggaactgtgtgggt 1969
|||||
Db 545 GCAGGGAAGACAAGAAAACACAAGAGAGCGGCTGTGGCCGGCTGGCACTGTGTGGGT 604
QY 1970 ttgtgtgatgccaactcggtgacctggtgtccctacgcaatagcagctgctgtggg 2029
|||||
Db 605 TTGTTGTGATGGCCACTCGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCTGTGGGG 664
QY 2030 aagaaggtgcgcccaagccagctgttctcccggaacaccagcagatccacacctgggca 2089
|||||
Db 665 AAGAAAGGCTGCCCAAGCCAGCTGTTCTCCCGGAGACACCAGCAGATCCACACCTGGGCA 724
QY 2090 cctcgtgttgtgtcttttttccctgtgtgaaagaagaagcaagcagaccctctc 2149
|||||
Db 725 CCTCCGTGTTGGTCTTTTTCCTTCCCTGTGTGAAGAAGAAACGGCACGACCCCTTCTC 784
QY 2150 aagctgtctactcaagacacatttggacaacccctggaagccatgccaagagagcgct 2209
|||||
Db 785 AAGCTGGCTCACTCAACACATTGGGACAAACCTTGACAGCCATGCCAGAGAGGCGCT 844
QY 2210 ttgaccggcccgagcgtaaagacacagagaanaatcaaatgtctcctactcagcgtgac 2269
|||||
Db 845 TTGACCGGCCCCAGAGCTTAAAGCACACAGAAATCAATGCTTCTACTCAGCGTGAC 904
QY 2270 ccaactttctagtgtgccaagcgcccccacacctcctcgtcagtaaccacacatcaccact 2329
|||||
Db 905 CCAACTTTTCTAGTGTGCCACGGGCCCCACACCTCCTGACGTACCCACACCATCACCACCT 964
QY 2330 gcttctctccaagatgcatctgtatcttctagtttcatatttcttcttggatgatg 2389
|||||
Db 965 GCTTCTCTTCCACAGTGTATCTGTATTCTTGTTCATTATTCTTTTGTGATGATAG 1024
QY 2390 aacatatataaatttcatlttgagaatttccaatgtatctagttaaatagcacaglt 2449
|||||
Db 1025 ACACTATATAAAATTTTCATTGTGAGAAATTTCTCAATTGTATCTAGTTAAATAGCACAGTT 1084
QY 2450 tggaaactgtctgagactgacttatacaataatctaaccgacaagaatcatatccatgt 2509
|||||
Db 1085 TGGAAACTGTGTGAGACTGACTTATATCAATATCTTAACCGACAAGATCATATCCATGT 1144
QY 2510 gtagtgtgttagacaatttatttcatltgactaaacccaggaaglttcaagtgtatgcaat 2569
|||||
Db 1145 GTATGTGTTAGACATTTTATTTCATTGTGACTTAACCCAGGACAGTTTCAGTGCATAAT 1204
QY 2570 tgtgtgccctctgtgtcagctgtaaacagltcctggaacttcaaaaaaccttgaataagltc 2629
|||||
Db 1205 TGTTGCCCTCTGTGTCAAGCTGAACAAGTCTCGACTTTCAAAAACCTTGAATAAGTCTC 1264
QY 2630 ccacagttgtataaatttggacaatttaggaattttaaactttagatgatacatltgtgtcc 2689
|||||
Db 1265 CCACAGTTGTATAAATGGACAATTAGGAATTTTAAACTTTAGATGATCATTTGGTTCC 1324
QY 2690 attttatttcaatttattatttltgttaatgcaaacaggaactttaaacttgaacttgaactct 2749
|||||
Db 1325 ATTTTATTTTCAATTTTATTTTGTTAATGCAACAGAGACTTAAATGAACTTTGATCTCT 1384
QY 2750 gttttaagaattataaaaaaacattgtgtatcctatacatatgctccttgaggactagct 2809
|||||
Db 1385 GTTTTAAAGATTATTAATAAACATTTGTGTATCTATACATATGGCTCTTGAGAGACTTAGCT 1444
QY 2810 ttcaactaacactacagatatgatctccatgtgtatccatatataaaccttgcaagatgtatttc 2869
|||||
Db 1445 TTCACTACACTACAGGATATGATCTCCATGTATGTCATATAAACCTTGACAGATGATTTTC 1504
QY 2870 cagaagtgtcgtatactgttaattacatctccatitagggtcgtgaaagaatgacctagctt 2929
|||||
Db 1505 CAGAGTGCTCGATACTGTTAATTACATCTCCAATTAGGGCTGAAGAAAGATGACCTACGTTT 1564
QY 2930 ctgtatacagctgtgtgtcttltgatgttgttactgtlacacagaagltgtgcaactga 2989

```
Db 1565 CTGTATACAGCTGTGTGCTTTGATGTTGTGTACTGTACACAGAGAGTGTGTCACTGA 1624
QY 2990 ggcctcgtcgtgtgtccgtatggaacacctgtagccctgcgagtttaagtactgtcca 3049
Db 1625 GGCCTCGCGTGTGGTCCGATGGAAGCCCTGGTAGCCCTGGCAGTTAAGTACTGCTTCCA 1684
QY 3050 ttcatgttttaacgctggaattttctcccatgtaagttaaaacttaagtgtt 3109
Db 1685 TTCAATGTTTACGCTGGAATTTTCTCCCATGGAATGTAAGTAAACCTTAAGTGTGT 1744
QY 3110 catcaataaatgtaataactaaaaa 3142
Db 1745 CATCAATAAATGTAATACTAAAAA 1777

RESULT 2
LOCUS BE867390 1018 bp mRNA EST 27-SEP-2000
DEFINITION 601442470F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846689 5',
mRNA sequence.
ACCESSION BE867390
VERSION BE867390.1 GI:10316166
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1018)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM546 row: 1 column: 18
High quality sequence stop: 654.
FEATURES
source location/Qualifiers
1..1018
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846689"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 295 a 267 c 262 g 194 t
ORIGIN
Query Match 22.3%; Score 701; DB 136; Length 1018;
Best Local Similarity 90.0%; Pred. No. 8.9e-162;
Matches 809; Conservative 0; Mismatches 80; Indels 10; Gaps 5;
```

```
Db 121 AAGAACGGGACACCAACAAGAGGTGGCCAGCGCAATGCAGCCGAGAACATCTGAGAGA 180
QY 1096 tccttggtttcaagtcccgacgcygcagccaccacaccccgactcaagtcagaggaga 1155
Db 181 TCCTTGGTTTCAAGAGTCCCGCAGCGCCAGCCACCAACCCGCACTCAAGTCAGAGAGA 240
QY 1156 agacacccataaagaaccagggatgtgaagaagaataacctttttgaacctgctctg 1215
Db 241 AGACACCCATTAAGAAACCCAGGGGATGGAAGAAAGTAACCTTTTGAACCTGCTCTG 300
QY 1216 gggatgaaaaatgggactagtaataaaggatgagttcaggatgcctatcctaagtcac 1275
Db 301 GGGATGAATAATGGACTAGTAATAAGAGGATGAGTTCAAGATGCCCTTATCTAAGTCATC 360
QY 1276 agcagctcgtcgtggaattctcccatggtgcccagggtcgccagcgtgaggagta 1335
Db 361 AGCAGCTGCCCTGCTGGAATTTCTCCCATGCTGCGCCGAGGTCGCCAGGCTGAGAGTTA 420
QY 1336 gtcaagacatcacaccaaagatttaccagggcagctccgaattcctgcgaagccacg 1395
Db 421 GTCACGACATCACACCAAAAGTTTACACAGGGCAGCTCCGAATCCTGCCAAGGCCACGG 480
QY 1396 taactgcatagtagcccgagagttgtgtatggygcacctgcgccacagccgagacca 1455
Db 481 TAACCTGCATGATAGCCCGAGAGTTGTGCTATGGGGGACCTCGCCACAGCCGAGACCA 540
QY 1456 tttaagaataacatctcttcaggccacgtlaccctcctcagagacctctcagagacct 1515
Db 541 TTTTAAAGATAACATCTCTTCAGGCCACAGTACCCCATGGAACCTCTCACAGACCTCTG 600
QY 1516 agcaactgactatcttccagagtcaggatccaggltgaatacaaaacttccca 1575
Db 601 AGCACTGACTATCTTCCAGAGTCCAGGATTCAGGTGAATACAAAGACTTCCCA 660
QY 1576 aaacaacaagaacgaatttgcattct--ctatacattgctcctcagccacctga 1632
Db 661 AAAATCACCAAGAAGCAATTCGATCTCTCTTATCAATGCGCTCCTCAGCCACTCTGAT 720
QY 1633 tcagccatggtatcgcgaagatgtgagtcctgcca--tgatatgctgcgctgaaca 1689
Db 721 CCAACATGTTATTCGCGCAGGTGTGGGAATCCTGGCAATGATCATGCGGTGGCTGACA 780
QY 1690 tcttaag-ttgcgtcgtgagttgaccacaaca-gtacagagatgccaagaacaggaa 1747
Db 781 TCTTTAAGATGGCTGCGCGAGGTGACCCACAAAGGTCAAAAAATTTCCCAAGAACGGGA 840
QY 1748 cggaccaatgctgtgtggtg--aggtgctgaacctttctggccatgaaccattata 1804
Db 841 GGGACCCAGTTGGTGTGGTGGCAGAGGCTCTGAACCATTTTGGGCTGAACCCATTAAA 899

RESULT 3
LOCUS BE889880 1036 bp mRNA EST 29-SEP-2000
DEFINITION 601512132F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3913590 5',
mRNA sequence.
ACCESSION BE889880
VERSION BE889880
KEYWORDS BE889880.1 GI:10347645
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1036)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1CM720 row: p column: 07
High quality sequence stop: 706.

FEATURES
source
1. .1036

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3913590"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 296 a 230 c 220 g 290 t
ORIGIN

Query Match 20.6%; Score 646.2; DB 137; Length 1036;
Best Local Similarity 91.5%; Pred. No. 2.8e-148;
Matches 740; Conservative 0; Mismatches 63; Indels 6; Gaps 5;

QY 2215 cggccccagagctaaagcaccagagaatcaatgcttcctactcagcgtgacccaac 2274
Db 1 CGGCCCCAGAGCTAAAGCACCCAGAGAAATCAATGCTTCCTACTCAGCGTGACCCAAC 60
QY 2275 tttctagtgccacggccccaccactctgcagtlaccacaccatcaccactgctt 2334
Db 61 TTTTCTAGTGTGCCACGGCCCCACCACCTCTCTGACGTACCCACACCATCACCACCTGCTTT 120
QY 2335 ctctccaacagtgatctgtatctctagttcatttctcttttgattgatalgacact 2394
Db 121 CTCTTCCAACAGATCTGTATTCTTAGTTTCATTATTTCTTTGATGTATGACACT 180
QY 2395 atataaatttccatttgaattctcgaattgtatctagttaaatagcacagtttgaa 2454
Db 181 ATATAAAATTTCATTGTGAGAAATTCGAATTGTATCTAGTTAAATAGCACAGTTTGAA 240
QY 2455 actgtctgagactgacttatacaataatcctaaccgacaaagatcatatccatgtgtatg 2514
Db 241 ACTTGTCTGAGACTGACTTTATCAATAATCTAACCGACAAGATCATATCCATGTGTATG 300
QY 2515 tggtagacattttatttattcatttgactaaccagagacagtttcagtgatgcaaatgtgt 2574
Db 301 TGGTTAGACATTTTATTTCATTGACTTAACCCAGACAGAGTTTCAGTGTGCAAAATTGTGT 360
QY 2575 gccctctggtcagctgaaacagtcctcgtgactttcaaaaaacctgaaatgaagtcctccaca 2634
Db 361 GCCCTCTGTTCAAGCTGAACAGTCCCTGGACTTTCAAAAACTTGAATGAGTCTCCACACA 420
QY 2635 gttgtataaattgacaattttagaattttaaactttagatgacattggttccattt 2694
Db 421 GTGTATAAATTGACAATTAGGAATTTTAACTTTAGATGATCATTTGGTTCCATT 480
QY 2695 tattcatttttattttgttaatgacaacagacttaaatgaacttgcactctgtttt 2754
Db 481 TATTTCAATTTTATTTTGTTAATGCAAAACGACTTAATGAACCTTGACTCTGTTT 540
QY 2755 aaagattataaaaaacattgtgtatctatataatgctccttgagacttagctttcac 2814
Db 541 AAAGATTATTAATAAACATGTGTATCTATACATATGGCTCTTGAGGACTTAGCTTTCA- 599
QY 2815 taactacagatatgatctcc-atgtagtcacataataaacctgcagagt--atttcca 2871
Db 600 TACACTACAGGATATGATCTCCAATGTAGTCCATATAAACCTGCAGAGTGGAATTTTCCAG 659
QY 2872 gagtgcctgatact-gttaaattacatctccatgagggctgaaaagaatgacctacgtttc 2930
Db 660 AAGTGTGATACTGTTTCATTACTCTCCATTAGGGTGAAAAAGAAATGAGACCTAGTTCTG 719

QY 2931 tgtatacagctgtgtgtgtcttttgatgtgtgttactgtlacacagaagtgtg-tgcactga 2989
Db 720 TATAAAGTGTGTGCTTGCCAGGTTGCGGGGTACTGCTGTAACACAGAGTGTGTCCCTGA 779
QY 2990 ggctctgcgtgtgtgtccgtatgaaaacc 3018
Db 780 GGGTCTGATGTGGCGCCGCTGTATACC 808

RESULT 4
AW952360 724 bp mRNA EST 01-JUN-2000
LOCUS
DEFINITION EST364430 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.
ACCESSION AW952360
VERSION AW952360.1 GI:8142042
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gasparid,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
COMMENT Unpublished (2000)
CONTACT: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@ligr.org
Plate: 38

FEATURES
source
1. .724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGB"
/note="Vector: pBluescriptSkm"

BASE COUNT 198 a 179 c 189 g 158 t
ORIGIN

Query Match 19.8%; Score 621.2; DB 97; Length 724;
Best Local Similarity 95.1%; Pred. No. 3.7e-142;
Matches 685; Conservative 0; Mismatches 28; Indels 7; Gaps 4;

QY 1375 cgaatcctcgccaagcgccaactgcatgatagcccgagagtggtgtatggygga 1434
Db 1 CGAATTCCTGCCAAGGCCACGGTAACCTGCCATGATAGCCCGAGAGTTGTTGATGGGGCA 60
QY 1435 cctcgccacagcccgagacattttaaagaatacatctctcagggccagtlaccatg 1494
Db 61 CCTGCCCCACAGCCGAGACCAATTTTAAAGAAATACATCTCTTCAGGCGCAGTACCCCATG 120
QY 1495 gacctctcagagacctctcgagcaactgacatcttccagagtcagggatccag 1554
Db 121 GACCTCTCACGAGACCCCTGAGCAACTGGACTATCTTCCAGAGTCCAGGATTCACAG 180
QY 1555 ttgaatacaaaagacttcccaaaaaacaacaagaagaatttgtatcttatcaattgct 1614
Db 181 TTGATATCAAAAGACTTCCCAAAAAACAACAAGAAGCAATTTGTATCTCTATCAATTGCT 240
QY 1615 cctctcagccacctctgatcagccatggtatcggaagatgttggagtcctgcacatgata 1674
Db 241 CCTCTCAGCCACACCTCTGATCAGCCATGTTATCGGCAAGGATGTGAGTCTCTGCCATGATA 300
QY 1675 tggctgcgtgaacatctttaaagttgctgtctgagttggaccaacaagaatgacagatgc 1734
Db 301 TGGCTGCGCTGAACATCTTAAAGTTGCTGTGTGAGTTGGACCAACAAGTACAGAGATGC 360

QY 1735 caagacaggaacgcgaccatgtctgtgtgtgaggtgctgnaacctttctgtgccaatg 1794
|||||
Db 361 CAAGAACAAGAAAGGACCAATGCTGTGTGTGGAGGTGCTGAACCTTTCTGGCCATG 420
QY 1795 aaccattataaaatcccaacatatatactgaaaaatactgaaactgtcttgaaaatttga 1854
|||||
Db 421 AACCATTTATTAATCCCAACATATATCTGAAATACTGAAACTGCTTTGAAATTTTGA 480
QY 1855 attctgatactcccaagtggcgcgagagacacggtggttaagagtgtggtgcagcgag 1914
|||||
Db 481 ATTTCTGATACCTCCAGTGGGCCGAGACACGGTGGGTAAAGATGTGGGCAGACGACG 540
QY 1915 ggaagacaacagaaacacaaagagcggtctgtgcccgg--ctggaactgtgtggtgtt 1972
|||||
Db 541 GGAAGACAACAGAAACACAAAGGAAGCGGCTGTGGCCGGCTTGACTGTGCGGGGCTTGTG 600
QY 1973 ttgtgatgccaactcgtgtgacctggtccctacgcaatagcagct--gcctgtgggga 2030
|||||
Db 601 TTGTGATGGCCCACTCGGGACCTGGCGGTCCCTACGCCAATACAAATTGGCTGTTGGGGA 660
QY 2031 agaaggtgcccagcc--agctgtgtctcccggtg--acaccagcagatccacacctgtg 2087
|||||
Db 661 AGAGGGGTGCCCAACCAAGCTGTTCTCCCGGGAACAACGCCGATTCCAAACTTGG 720

RESULT 5
BE894295 752 bp mRNA EST 29-SEP-2000
LOCUS BE894295 601437666F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922529 5',
DEFINITION mRNA sequence.
ACCESSION BE894295
VERSION BE894295.1 GI:10356518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DCFD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM744 row: d column: 18
High quality sequence stop: 654.

FEATURES
source
1. 752
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922529"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 223 a 194 c 192 g 143 t
ORIGIN

Query Match 19.6%; Score 617.2; DB 137; Length 752;
Best Local Similarity 94.2%; Pred. No. 3.6e-141;
Matches 684; Conservative 0; Mismatches 38; Indels 4; Gaps 4;

QY 883 aatatgcccagggatcaatccgatlttagccgacttgcccagatccagcgcaaaaaag 942
|||||
Db 1 AATATGGCCAGGGGATCAATCCGATTAACCGACTGGCCAGATCCAGCAGGCAAAAAAG 60
QY 943 agaaggaagccgaagtlacacgcctcctcacagaagcgagcgctcccgccgcgcaaggagt 1002
|||||
Db 61 AGAAGAGCCAGAGTACACGCTCCTCACAGAGCGAGGCGCTCCCGCGCCGAGGAGTTTG 120
QY 1003 tgatgcaggtgaaggtltggaaccacactgcagaaggaagcgagcaccacaagaaggttg 1062
|||||
Db 121 TGATGCAGGTGAAGGTTGGAANAACCACTGCAGAAGGAACGGGCACCAACAAGAAAGTGG 180
QY 1063 ccaagcgcaatgcagccgagacacatgctgagatccttglttcaagltcccgcaagcgc 1122
|||||
Db 181 CCAAGCGCAATGCAGCCGAGAACATGCTGAGATCCTTGTTCAAAAGTCCCGCAGGCGC 240
QY 1123 agccaccaaaccgcactcaagtcaagaggaagacacccaataaagacagggatg 1182
|||||
Db 241 AGCCCAACAACCCCGCACTCAAGTCAGAGGAGAAGACACACCATAAAGAACCA -GGGATG 299
QY 1183 gaagaaaagtaacctttttgaacctgtctgtggatgaaatggyactagtaataag 1242
|||||
Db 300 GAAGAAAAGTAACCTTTTGTGAACCTGCTCTGGGATGAANAATGGGACTAGTAATAAG 359
QY 1243 aggatgagtlcaggtatgccttatcttaagltcacagcagctgctgtgaaatcttccca 1302
|||||
Db 360 AGGATGAGTTCAGGATGCTTATCTTAAGTCATCAGCAGCTGCTGCTGGAATTTCCCA 419
QY 1303 tgggtccgaggtgcgccaggtctgtagagtltagtcaagacatcacaccaagaattta 1362
|||||
Db 420 TGGTGCCCGAGGTGCGCCAGGCTGTAGAGATTAGTCAAGGACATCACACCAAGATTTTA 479
QY 1363 ccagggcagctccgaatccttgccaaagccacggtacgttaactgcatatagcccgaggtgt 1422
|||||
Db 480 CCAGGACAGCTCCGAATCCTGCAAGGCCACCGTAACCTGCATGATAGCCCGAGAGTTGT 539
QY 1423 tgtatggggcacctcgcgccacagccgagacattttaa--gaatacatctctcagggc 1481
|||||
Db 540 AGTATGGGGGCACCTCGCCACAGCCGAGATCATTTTAAACGAATAACATCTCTCAGGC 599
QY 1482 caagtaccccatgacctctcagagagaccctctgagcaactggaactatcttccaagatc 1541
|||||
Db 600 CACGTACCCCATGGA-CTCTACGAGATCTCTTGAGCAA-TGGACTATCTTCCAGAGTTC 657
QY 1542 caggatccaggtltgaatacaagaacttccccaaaaacaagaagaattgtatct 1601
|||||
Db 658 CAGGATTCGGGTTGATACAAAGACTTTCCCAAAAAACGAAGAATTGTTCTCTATCA 717
QY 1602 ctatc 1607
|||||
Db 718 ATGCTC 723

RESULT 6
BE379687/c 821 bp mRNA EST 21-JUL-2000
LOCUS BE379687 601159257T1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510806 3',
DEFINITION mRNA sequence.
ACCESSION BE379687
VERSION BE379687.1 GI:9325052
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 821)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov


```
QY 2309 agtaccacacacatcaccaactgtcttctctccaacagtgatctgtatcttcttagttcat 2368
      |||||||
Db 121 AGTACCACACACATCACCACTGCTTCTCTCCACACAGTGTCTGTATCTTAGTTTCAT 180
QY 2369 tatcttcttcttgatgtatgacactatataaaatttcatcttgagaatttctcaattgt 2428
      |||||||
Db 181 TATTTTCTTTTGATGTATGACACTATATAAAATTTTCATTTGAGAAATTTCTCAATGT 240
QY 2429 atctagtttaaatagcacagtttggaaactgtctgagactgacttatacaataatcaac 2488
      |||||||
Db 241 ATCTAGTTAATAGCACAGTTTGAAACTGTGTGAGACTGACTTTATCAATAATCTAAC 300
QY 2489 cgacaaagatcatatccatgtgtatgtgttagacatttattcatctgactaaccag 2548
      |||||||
Db 301 CGACAAGATCATATCCATGTGTATGTGTGATGACATTTTATTTTCATGACTAACCCAG 360
QY 2549 gacagtttcagtgatgcaaatgtgtgcccctcgtgttcagctgaaacagtcctgactt 2608
      |||||||
Db 361 GACAGTTTCAGTGATGCAAAATGTGTGCCCTCTGTCTCAGCTGAACAGTCTGGACTTT 420
QY 2609 caaaaacctgaataagctctccacacagttgtataaattggacaatttgaattttaa 2668
      |||||||
Db 421 CAAAAACCTTGAATAAGTCTCCACAGTTGTATTAATTGACAAATTTAGGAATTTTAAAC 480
QY 2669 tttagatgatcatctgtgtccatttatttatttatttatttatttatttatttatttatt 2728
      |||||||
Db 481 TTTAGATGATCATTTGGTTCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 540
QY 2729 cttaaatgaacttgcctcctgttttaaagattataaaacacattgtgtatctatcat 2788
      |||||||
Db 541 CTTAAATGAACCTTGTACTCTGTCTTTTAAAGATTTTAAACATTTGTATCTATACAT 600
QY 2789 atgg 2792
      |||||
Db 601 ATGG 604

RESULT 8
LOCUS BE613972 686 bp mRNA EST 24-AUG-2000
DEFINITION 601504093F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5',
      mRNA sequence.
ACCESSION BE613972
VERSION BE613972.1 GI:9895569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 686)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
      Tel: (301) 496-1550
      Email: Robert_Strausberg@nih.gov
      Tissue Procurement: ATCC
      cDNA Library Preparation: Life Technologies, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LICM700 row: j column: 24
      High quality sequence stop: 675.
FEATURES
      Location/Qualifiers
          1..686
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:3905783"
              /clone_lib="NIH_MGC_71"
              /tissue_type="leiomyosarcoma"
```

```
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
      Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
      Average insert size 2.1 kb."
BASE COUNT 226 a 156 c 196 g 108 t
ORIGIN
Query Match 19.2%; Score 603.8; DB 110; Length 686;
Best Local Similarity 97.1%; Pred. No. 7e-138;
Matches 669; Conservative 0; Mismatches 12; Indels 8; Gaps 5;

QY 490 aacacgatgctgtgccaaagcgttgagatcctgcagaatgagccctgcagagagc 549
      |||||||
Db 1 AACACGATGCTGCTGCCAAAGCCTTGAGATCTGCAGATGAGCCCTGCCAGAGAGGC 60
QY 550 tggaggtgaatggaaggaatccgaagaagaatctcaataatctgaataagltcaag 609
      |||||||
Db 61 TGGAGGTGAATGGAAGAGAAATCCGAAGAAGAAATCTCAATAAATCTGAATAAAGTCAAG 120
QY 610 tgtttgagatgtgcaactaaacggaacttgcctgtgaatttcgaggtgcccggagagt 669
      |||||||
Db 121 TGTTTGAGATTGCACCTTAAACGGAACCTTGCCGTGATTTTCGAGGTGCCCCGGAGAGTG 180
QY 670 gcccacccacatgaagaacttgtgaccaaagtctggttgggaggttgttggggaag 729
      |||||||
Db 181 GCCCACCCACATGAAGAAC-TTGTGACCAGGTTTCGGT--GGGGAGTTGGGGGAAG 237
QY 730 gtgaaggggaaagcaagaagatttcaaagaanaatgccgcataagctgttcttgaggagc 789
      |||||||
Db 238 GTGAAGGGGAAAGCAAGAGATTTCAAAGAAATATGCCCATAGCTGTCTTGAGGAGC 297
QY 790 tgaagaagtlaccgcccctgtcgtcagltgaacgaglaaagcctagaatcaaaaaaaga 849
      |||||||
Db 298 TGAAGAAGTTACCGCCCTGCTGCTGAGTTGAACGAGTAAAGCCTAGAAATCAAAAAAGAAA 357
QY 850 caaaaccatagtcagcagcacagacaagcccagatalgycaggygatcaatccgatta 909
      |||||||
Db 358 CAAAACCCATAGTCAAGCCACAGACAAGCCCAAGATATGGCCA-GGGAATCAATCCGATTA 416
QY 910 gccgactggcccagatccagcagcgaagaaagagagagccagagtcacgctctca 969
      |||||||
Db 417 GCCGACTGGCCCCAGATCCAGCAGGCAAAAGAGAGAGACGACAGATACGCTCTCA 476
QY 970 cagagcgagcgctcccgccgcccagggaggttctgtatgcaggtgaaggttggaaccaca 1029
      |||||||
Db 477 CAGAGCGAGGCGCTCCCGCCCGCAGGGAGTTGTGATGCAGGTGAAGTTGGAACACACA 536
QY 1030 ctgcagaaggaacgggcaaccaagaaggtggtccaaagcgcaatgcagccgagataatgc 1089
      |||||||
Db 537 CTGCAGAAAGAAACGGGCACCAACAAGAGGTGTGCCAAGCGCAATGCAGCCGAGAACATGC 596
QY 1090 ttgagatccttg-tttcaaagtcctcgagcggcagcccaaccacccgactcaagtca 1148
      |||||||
Db 597 TGGAGATCCTTGTGTTTCAAAAGTCCCGCAGGCGCAGCCACCAAA--CCGACTCAAGTCA 653
QY 1149 gaggaagaagacaccataaagaacagg 1177
      |||||||
Db 654 GAGGAGAAGACACCCCATTAAGAAACAGGG 682

RESULT 9
LOCUS AV707406 674 bp mRNA EST 09-OCT-2000
DEFINITION AV707406 ADB Homo sapiens cDNA clone ADBCMG03 5', mRNA sequence.
ACCESSION AV707406
VERSION AV707406.1 GI:10724671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 674)
```


QY 258 acccctactgtagaactaaatgcactgtgcatgaaacttggaaaaaaccaatgtataag 317
|||||
Db 564 ACCCCTACTGTAGAACTAAATGCACCTGTGCATGAACCTTGAAAAAACCAATGTATAAG 505
QY 318 cctgtgacccttactctcgatgcagtcaccctataactacaacatgagaggtgct 377
|||||
Db 504 CCTGTGACCCCTTACTCTCGAGTGCAGTCCACCTATACTACACATGAGAGAGGTGCT 445
QY 378 tatcccccaggtactcttaccattccagttccaccttacttatacaagtgaactt 437
|||||
Db 444 TATCCCCGAGGTACTTTTACCCATTTCACAGTCCACCTTACTTTATCAAGTGAACCTT 385
QY 438 tctgtgggagagcagcaatttaattgcaaaaggaagacaagcagctgcgaaaaacagat 497
|||||
Db 384 TCTGTGGGAGGACAGCAATTTAATGCCAAAGGAAGACAGACAGGCTGCCAAACACGAT 325
QY 498 gctgtgtccaaagcgltgagatcctgcagaatgagccccctgcagagagagctgaggtg 557
|||||
Db 324 GCTGCTGCCAAAGCGTTGAGATCCTGCAGAAATGAGCCCCCTGCCAGAGAGGCTGGAGGTG 265
QY 558 aatggaagagaatccgaaagaaaatctcaataatctgaataaagtcagtglttgaag 617
|||||
Db 264 AATGGAAGAGAATCCGAAGAGAAAAATCTCAATAAATCTGAATAAAGTCAAGTGTTTGAG 205
QY 618 attgcaacttaacggaacttgctgtgaatttcgaggtgccccggagagtgccccacc 677
|||||
Db 204 ATTGCACCTTAAACGGAACCTTGCTGTGAATTTCCAGGTGGCCCCGGAGAGTGGCCCCACCC 145
QY 678 cacatgaagaactctgtgaccaaagtttcggttgaggaggttgtgagggaaggtgaagg 737
|||||
Db 144 CACATGAAGAAGACTTGTGACCAAGGTTTCGGTGGGAGTTTGTGGGGGAAGGTGAAGGG 85
QY 738 aaaaagcaagaagatttcaagaaaaaatgccgcacatagctgttcttgaggagctgaaga 795
|||||
Db 84 AAAAGCAAGAAGATTTCAAAGAAAAAATGCCGCCATAGCTGTCTTGAGAGGAGCTGAAGA 27

RESULT 11

LOCUS BE674157 616 bp mRNA EST 08-SEP-2000
DEFINITION 7d76a04.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3278862 3'
similar to TR:095793 O95793 STAUFPN PROTEIN. ; mRNA sequence.
ACCESSION BE674157
VERSION BE674157.1 GI:10034698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 616)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 469.
Location/Qualifiers
1. 616

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3278862"
/clone_lib="NCI-CGAP_Lu24"

/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

Query Match 18.9%; Score 594.4; DB 111; Length 616;
Best Local Similarity 98.4%; Pred. No. 1.4e-135;
Matches 598; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1155 aagacaccataaagaaccaggagatggaagaagaatgaacctttttgaactgtgctt 1214
|||||
Db 8 AAGACACCCATTAAGAAACCAAGGGGATGGAAGAAAGTAACCTTTTGAACCTGGCTCT 67
QY 1215 ggggatgaaaaatgagactgtaataaagagatgagttcagatgacctatctaagtc 1274
|||||
Db 68 GGGGATGAATAATGGAGACTAGTAATAAGAGAGATGATTCAGATGCTTATCTAAGTCAT 127
QY 1275 cagcagctgcctgtcgtgaattctccatggtgcccgaggtgcgccagctgtagagtt 1334
|||||
Db 128 CAGCAGCTGCCTGTGGAATCTTCCCATGTGCCCCGAGGTGCGCCAGGCTGTAGAGATT 187
QY 1335 agtcaaggacatcacaccaagaatttaccaggagcagctccgaatctctgccaaggccacg 1394
|||||
Db 188 AGTCAAGGACATCACACCAAGATTTTACCAGGCGCAGCTCCGAATCCTGCAAGGCCACG 247
QY 1395 gtaactgcatgatagcccgagagttgtgtatgagggggcaacctcgccacagccgagacc 1454
|||||
Db 248 GTAACCTGCCATGATAGCCCGAGAGAGTTGTGTATGGGGGCACCTCGGCCACAGCCGAGACC 307
QY 1455 attttaagaataacatctcttccagggccagctaccatgtgacctctcaagagacctct 1514
|||||
Db 308 ATTTTAAGAATAACATCTCTTCCAGGCCACGATGCCCATGAGACCTCTCAGAGACCTCT 367
QY 1515 gagcaactgactatcttccagaggtccaggagatccaggttgtaataaagaacttcccc 1574
|||||
Db 368 GAGCAACTGACTATCTTCCAGAGTCCAGGAGGATTCAGGTTGAATACAAAGACTTCCCC 427
QY 1575 aaaaacaacaagaacgaattgtatctctatcaaatgtctcctccagccacctgtgac 1634
|||||
Db 428 AAAAACAACAAGAAGAAATTGTATCTCTTATCAATTGCTCCTTCANCCACCTCTGATC 487
QY 1635 agccatgatatcggaagatgtgagtgctgcctgatgatagtgctgcgtgaacatctta 1694
|||||
Db 488 AGCCATGGTATCAGCAAGATGTGAGTCTGCCATGATATGGCTGCGCTGAACATCTTA 547
QY 1695 aagttgctgtctgagttgaccacaacaagtacagagatgccaagaagaaacgagacca 1754
|||||
Db 548 GAGTTGCTGTCTGAATGTGACCACAACAANTACAGAGATGCCAAGAACANAGAGACCA 607
QY 1755 atgtctgt 1762
|||||
Db 608 ATGTCTGT 615

RESULT 12

LOCUS AW162002/c 594 bp mRNA EST 09-NOV-1999
DEFINITION au72c03.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2781796 3' similar to TR:095793 O95793 STAUFPN PROTEIN. ;
mRNA sequence.
ACCESSION AW162002
VERSION AW162002.1 GI:6301035
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 594)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maira,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: au72c03.y1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 464.
FEATURES
source
1..594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2781796"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene); Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was prepared from human fetal brain tissue. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-GAGAGAGAGAGAGAGCTCAAGATCCTTAATTAATTATCCCCCCCCCC-3' and 3' adaptor sequence: 5'-GAGAGAGAGAGAGCTCGATTGTTTCTTTT-3'. The library was size-selected for >0.5 kb inserts and has an average insert size estimated at 1.2 kb. This library was constructed using the CAP-trapper method for full-length enrichment and has not undergone amplification. Library was constructed by Dr. Claudio Schneider (LNCIB-Area Science Park, Trieste, Italy)."
BASE COUNT 130 a 152 c 123 g 188 t 1 others
ORIGIN
Query Match 18.7%; Score 588.2; DB 40; Length 594;
Best Local Similarity 99.3%; Pred. No. 4.7e-134;
Matches 590; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 249 gaaagcataaccctactgtagaactaaatgacactgtgcatgaaacttgaaacaa 308
DB 594 GAAAGCATTAACCCCTACTGTAGACTANATGCACTGTGCATGAACCTTGAAAAACCA 535
QY 309 atgtataagcctgtgacccctactctcgatgacgtccacctataactacaacatgaga 368
DB 534 ATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATACTACAACATGAGA 475
QY 369 ggaggtgtctatcccccgaggtactttaccattccagttccaccttactatcaa 428
DB 474 GGAGGTGCTATCCCCCGAGGTACTTTTACCCATTTCCAGTTCCACCTTACTTTATCAA 415
QY 429 gtggaacttctgtgaggagacagaatttaatgccaaggaagacaagacagctgag 488
DB 414 GTGGAACCTTCTGTGGGAGACAGCAATTTAATGGCAAGAAAGACAAGACAGGCTGCG 355
QY 489 aaacacgatgctgctgcacaaagcgttgagagatcctgcagaaatgagccctgcagagag 548
DB 354 AAACACGATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGATGAGCCCTGCAGAGAGG 295

QY 549 ctgaggtgaatgyaagaatccgaagaagaatctcaataaatctgaataatgaatcaa 608
DB 294 CTGAGGTGAATGGAAGAAGATCCGAGAGAAATCTCAATAATCTGAATAATAGTCAA 235
QY 609 gtgtttgagattgcacttaaacggaactgacctgtgaatttcogaggtgccccggagagt 668
DB 234 GTGTTTGAGATTGCACCTTAACGGAAGTTGCTGTGAATTCGAGGTGGCCCGGAGAGT 175
QY 669 ggcacacacatgaagaacttctgaccaaaggttcggttggggaggttctgaggagaa 728
DB 174 GGCCACCCACATGAGAACTTTGTGACCAAGGTTTCGGTTGGGAGTTGTGGGGGAA 115
QY 729 ggtgaaggaagaaagcaagaatttcaagaaatgcccacatagctgtcttgaggag 788
DB 114 GGTGAAGGGAAGCAAGAGATTTCAGAAAGAAATGCCGCATAGCTGTTCTTGAGAG 55
QY 789 ctgaagaagtaccgccccctgctgcagtgaacgagtaagcctagaatcaa 842
DB 54 CTGAAGAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAGCCTAGAAAAAAA 1
RESULT 13
AV695897 612 bp mRNA EST 25-SEP-2000
LOCUS AV695897 GKC Homo sapiens cDNA clone GKCCEF08 5', mRNA sequence.
DEFINITION AV695897
ACCESSION AV695897
VERSION AV695897.1 GI:10297760
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 612)
AUTHORS Wu,T., Qian,B., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GK- clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCEF08"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 149 a 164 c 125 g 174 t
ORIGIN
Query Match 18.2%; Score 571.4; DB 38; Length 612;
Best Local Similarity 98.4%; Pred. No. 6.5e-130;
Matches 611; Conservative 0; Mismatches 1; Indels 9; Gaps 3;
QY 1988 ggtacactgagcgtccctacgaatagacagctgctgtggggaagagggctgccagcc 2047
DB 1 GGTGACCTGGCGGCTCCCTACGCAATATAGCAGCTGCTGTGGGAGAAAGGCTGCCAGCC 60
QY 2048 agctgttctcccgaggacaccagcagatccacaccctgggcacctccgtgtgtcttt 2107

Db 61 AGCTGGTCTCCGGGACACAGCAGATCCACACCTGGGCACCTCCGTGTTGTTTT 120

QY 2108 ttcttccctgtgtgtaagaagaacagcgacgacccctctcaagtgtcactcagac 2167
|||||
Db 121 TTTTCCCTGTGTGAAGAAGAACGCGACGACCCCTTCTCAAGCTGGCTCAGACTCAGAC 180

QY 2168 acattggacaacacctggacagccatgcccagagagagcccttgcgccccagagct 2227
|||||
Db 181 ACATTGGGACAACCCCTGGACAGCCATGCCAGAGAGGCCCTTGACCGGCCCCAGAGCT 240

QY 2228 aaaagcaccagagaaatcaaatgcttcctactcagcgtgacccaactttctagtgtc 2287
|||||
Db 241 AAAGCACACAGAGAATCAATGCTTCTTACAGCGTGACCCAACTTTCTAGTGTGC 300

QY 2288 cacygccccaccacctcctgtagtaccacacatacaccactgtcttctctccaagct 2347
|||||
Db 301 CACGGCCCCACACCTCTCCAGTACCCACACCATCACCACCTGCTTCTTCCAACAGT 360

QY 2348 gactgtattcttagttcattatttcttctgattgatatgacacctataaaatttc 2407
|||||
Db 361 GATCTGTATTCTTAGTTTCAATTAATTTCTTTGATTGATATGACACTAATAT---TTTTC 416

QY 2408 atttgaagaattctcaattgtatctagttaaatagcacagtttgaaactgtctgagac 2467
|||||
Db 417 ATTTGAGAATTTCTCAATTGTATCTAGTTAAATAGCACAGTTTGAACCTTGCTGAGAC 476

QY 2468 tgactttatcaataatcaaccgacaagaatcatatcatgtgtatgtgttagacattt 2527
|||||
Db 477 TGACTTTATCAATATCAATCAACCGACAAGATCATATCCATGTGATGTGTTAGACATTT 536

QY 2528 ttatttcattgactaaccagagacagtttcagtgtatgcaaatgtgtgccctctgttca 2587
|||||
Db 537 TTATTTTCATTGACTAACCCAGAGACAGTTTCAGTATGCG--ATTGTGTGCCCTCTGTTCA 594

QY 2588 gctgaacacagtcctggaactt 2608
|||||
Db 595 GCTG---CAGTCTGACTTT 612

RESULT 14

BE617223 749 bp mRNA EST 24-AUG-2000

LOCUS 60141906F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846265 5',
DEFINITION mRNA sequence.

ACCESSION BE617223

VERSION BE617223.1 GI:9888161

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 749)

AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM545 row: k column: 02
High quality sequence stop: 705.

FEATURES

source 1..749

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846265"
/clone_lib="NIH_MGC_65"

/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 244 a 166 c 223 g 116 t
ORIGIN

Query Match 18.0%; Score 565.4; DB 110; Length 749;
Best Local Similarity 95.1%; Pred. No. 2.1e-128;
Matches 714; Conservative 0; Mismatches 21; Indels 16; Gaps 12;

QY 471 aagacaagacagcgtgcgaaacaacgatgctgtgccaacaa-gcgttgagatcctgcagaa 529
|||||
Db 1 AAGACAAGACAGCGCTGCCAACAACGATGCTGCTGCCAAAGCGCTTGAGATCCTGCAGAA 60

QY 530 tgaagccctgcagagagcgtgaggtgaatgaaagaaatccgaagaaagaaatctcaa 589
|||||
Db 61 TGAGCCCCCTGCCAGAGAGGCTGAGGTGAATGGAAGAAATCCGAAGAAAGAAATCTCAA 120

QY 590 taaatctgaataaagtcgaagtgtt-gagatgcaacttaacggaactgtcgtgaatt 648
|||||
Db 121 TAAATCTGAATTAAGTCAAGTGTGAGATGCACTTAAACGGAACCTTGCTGTGAATT 180

QY 649 tggaggtgccccgagagtggtgccccacatagaagaacttgtgaccaagtttcgg 708
|||||
Db 181 TCGAGGTGGCCCCGGGAGAGTGGCCCCACCCACATGAAGAAC-TTGTTGACCAAGTTTCGG 239

QY 709 ttggygagtttgttgyggygaagtgaaggyaaagcaagaatttcaaagaaatgtccg 768
|||||
Db 240 TTGGGGAG-TTGTGGGGGAAGGTGAAGGGAAGCAAGAAGATTTCAGAAAGAAATGCCG 298

QY 769 ccatagtcttcttgagagagctgaagaagtlaccgccccctgacctgcaagttgaacgata 828
|||||
Db 299 CCATAGCTGTTCTTGAGAGAGCTGAAGAAGTTACCG-CCCTGCTGCAGTTGAACGAGTAA 357

QY 829 agcctagaatcaaaaaaagaaaaaaccatagtcgaagccacagacaagccagaatatg 888
|||||
Db 358 AGCCTAGAATCAAAAAGAAACAAACCCATAGTCAAGCCACAGACAGACCCAGATATG 417

QY 889 gccagggatcaatccgatttagccgactgtgcccagatccagcagcgcaaaaaagagaag 948
|||||
Db 418 GCCA-GGGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGGCAAAAAAGAGAAG 476

QY 949 agccagagtacacgctcctcacaagagcgagcctcccgcgccgagggagttgtatgc 1008
|||||
Db 477 AGCCAGAGTACACGCTCCTCACAGAGCGAGG-CTCCCGCGCCGACGGAG-TTGTGATGC 534

QY 1009 aggtgaaggttgaaacacacactgcagaagaagcggcaccac-aagaaggtgccaag 1067
|||||
Db 535 AGGTGAAGGTGGGAACCACTGTCAGAGAAGAACGGGCACCAACAAGAGGTGGCCAAG 594

QY 1068 cgcaatgcagccgagacaatgctgagatcctgtgtt--caaaagtcgagcgagcagc 1125
|||||
Db 595 CGCAATGCAGCCGAGACATGCTGGAGATCTTGTGTTTCAAAAAGTCCCGCAGGGCAGC 654

QY 1126 ccaccaaacccgactcactcaagtcagagagagacaccataaagaacaggagatgaa 1185
|||||
Db 655 CAACAAACCGGAGCTC---AGTCAGAGAGAGAGACACCCATAA--GAACCGGGGATGAA 709

QY 1186 gaaaagtaacctttttgaacctgtgtg 1216
|||||
Db 710 GACACGTAACCTTTTGGAACTGGGCGCGGG 740

RESULT 15

BE810330 609 bp mRNA EST 21-SEP-2000

LOCUS BE810330

DEFINITION MR4-PT0052-220500-206-f06 PT0052 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE810330

VERSION BE810330.1 GI:10242524

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 609)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4&t2=MR4-PT0052-220500-206-f06&t3=2000-05-22&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 589.
Location/Qualifiers
1. 609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="PT0052"
/dev_stage="Adult"
/note="Organ: pnet; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 173 a 153 c 145 g 138 t
ORIGIN
Query Match 17.7%; Score 555; DB 136; Length 609;
Best Local Similarity 97.9%; Pred. No. 7.1e-126;
Matches 594; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
QY 1222 aaaatggactagtaataaaggatgagtcagatgacctatctaatgcatcagcagc 1281
DB 6 AACATGGGACTAGTAAT-AAGAGGATGAG-TCAGGATGCTTATCTAAGTCATCAGCAGC 63
QY 1282 tggcctgtggaatcttcccatggtgccgaagtcgcccaggctgtagagtaagtaag 1341
DB 64 TGCCCTGCT-GAATTCTTCCCATGTGCGCCGAGGTGCCCCAGGCTGTAGAGTTAGTCAAG 122
QY 1342 gacatcacaccaagaattttacaggcgagctccgaatcctgccaagggccagcytaactg 1401
DB 123 GACATCACACCAAGAATTTTACAGGGGAGCTCCGAATCTGCGCAGGGCCACGGTAACTG 182
QY 1402 ccatgtagcccgagagtggtgtatggtgggcaacctcgcccacagccgagaccatttta 1461
DB 183 CCATGATAGCCCGAGAGTAGTGTATGGGGGCACCTCGCCACAGCCGAGACCATTTTAA 242
QY 1462 agaataacatctcttcagggccagctaccccatgagaccttcacagagagacctctagcaac 1521
DB 243 AGAATAACATCTCTTCAGGGCCAGCTAACCCATGAGACCTTTCACGAGACCCCTCTGAGCAAC 302
QY 1522 tggacctcttccagagtgccaggagttccaggttgaaatacaagaacttccccaaaaaca 1581

DB 303 TGGACTATCTTTCCAGAGTCCAGGGATTCAGGTTGAATACAAAGACTTCCCAAAAACA 362
QY 1582 acaagaacgaatttgtatctcttatacaattgtctccttcagccaccctctgatacagcatg 1641
DB 363 ACAAGACGAATTGTGTATCTTATCAATGAGCTCCCTCCAGCCACCTGTGATCAGCCATG 422
QY 1642 gtatcggaagatgtgtgagtccttcgcatgatgatgtgctgcgtgaacatcttaagtgtgc 1701
DB 423 GTATCGGCAAGGATGTGAGTCTGCCATGATATGCGCTGCCGTGAACATCTTAAAGTTGC 482
QY 1702 tgtctgagttggaccaacaagtaacagatgccaagaacaggaaacggaccaatgltctg 1761
DB 483 TGTCGTGAGTAGGACCACCAAGTACAGAGATGCCAAGAACGGAACGACCAATGTCTG 542
QY 1762 tgtgtggaggtgtgtaaccttctctggtgcatgaaccattataaaatcccaatatata 1821
DB 543 TGTGTGGAGGTGCTGAACCTTTTCTGGGCATGAACCAATTATAAAATCTCAACATATATA 602
QY 1822 ctgaaaa 1828
DB 603 CTGGCAA 609

Search completed: April 5, 2001, 00:29:19
Job time: 53055 sec

OM of: US-09-316-048-6 to: GenEmbl:* out_format : pfs
Date: Apr 5, 2001 10:32 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09316048/runat_01042001_170039_7742/app_query.fasta_1.1293
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09316048@cgn1_1_9054 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-316-048-6
Query length: 577
Database: GenEmbl:*
Database sequences: 1118133
Database length: -1736092196
Search time (sec): 4765.960000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
gb_pr3:AF061939	+ 2970.00	3157.73	9.9e-168	3506	! AF061939 Homo sapiens stauflen
gb_pr3:AF061940	+ 2664.50	2833.03	1.2e-149	3142	! AF061940 Homo sapiens stauflen
gb_pr3:AF061938	+ 2659.00	2826.95	1.2e-149	3217	! AF061938 Homo sapiens stauflen
gb_pr3:AF061941	+ 2649.00	2815.92	1.1e-148	3349	! AF061941 Homo sapiens stauflen
gb_pr4:AK025519	+ 2587.50	2750.78	4.6e-145	3194	! AK025519 Homo sapiens CDNA: FL
gb_pr6:HS132258	+ 2577.50	2740.50	1.7e-144	3066	! AJ132258 Homo sapiens mRNA for
gb_ro:AF290989	+ 2445.50	2599.84	1.2e-136	3041	! AF290989 Rattus norvegicus sta
gb_ro:AF227200	+ 2432.50	2585.93	7.0e-136	3059	! AF227200 Rattus norvegicus sta
gb_ro:AF061942	+ 2422.50	2575.88	2.5e-135	2860	! AF061942 Mus musculus RNA-bind
gb_pr6:HS19062	+ 1081.50	1142.98	1.7e-55	4058	! Y19062 Homo sapiens mRNA for
gb_pr4:AK002152	+ 1067.50	1130.92	7.8e-55	2968	! AK002152 Homo sapiens CDNA FL
gb_htg14:AC063967	- 921.00	935.10	6.3e-44	225016	! AC063967 Mus musculus clone
gb_pr6:HSB140E4	+ 798.00	806.90	8.7e-37	163253	! AL121877 Human DNA sequence
gb_pr4:AL133174	- 753.00	762.22	2.7e-34	113836	! AL133174 Human DNA sequence
gb_in3:DROSTAUFEN	+ 716.00	750.97	4.8e-30	5237	! M69111 D.melanogaster mRNA, co
gb_in1:AC004336	+ 676.00	685.80	1.1e-33	61267	! AC004336 Drosophila melanoga
gb_in1:AE003800	- 676.00	672.32	2.7e-29	267271	! AE003800 Drosophila melanoga
gb_htg6:AC017563	- 674.00	677.47	1.4e-29	120625	! AC017563 Drosophila melanoga
gb_in2:AF225925	+ 656.00	683.12	6.8e-30	8002	! AF225925 Musca domestica stau
gb_in2:AF225924	+ 641.50	667.20	5.2e-29	8409	! AF225924 Drosophila virilis st
gb_pr4:AK001576	+ 581.00	611.63	6.5e-26	3173	! AK001576 Homo sapiens CDNA FL
gb_ro:RNO010200	+ 571.00	620.89	2.0e-26	360	! AJ010200 Rattus norvegicus mRNA
gb_pr4:AK023314	+ 462.00	488.94	4.4e-19	2008	! AK023314 Homo sapiens CDNA FL
gb_htg14:AC055879	- 411.00	394.43	8.2e-14	160979	! AC055879 Homo sapiens chromo
gb_htg3:AC011492	- 304.50	281.07	1.7e-07	157803	! AC011492 Homo sapiens chromo
gb_htg15:AC068845	- 301.00	275.43	3.5e-07	194407	! AC068845 Homo sapiens chromo
gb_pr7:HS100002	+ 271.00	290.29	5.2e-08	1166	! AL078636 Homo sapiens mRNA fu
gb_in2:CELF5A4	- 270.50	256.92	3.7e-06	42099	! U67949 Caenorhabditis elegans
gb_htg14:AC055879	- 264.00	237.71	4.4e-05	160979	! AC055879 Homo sapiens chromo
gb_htg22:AL390198	- 254.00	220.40	0.0004	332811	! AL390198 Homo sapiens chromo
gb_htg12:AC027018	- 250.00	222.20	0.0003	171591	! AC027018 Homo sapiens chromo
gb_pr8:HUMTRBP	+ 241.00	256.84	3.8e-06	1368	! M60801 Human TAR RNA binding p
gb_pr7:HSU08998	+ 239.50	254.61	5.0e-06	1465	! U08998 Homo sapiens TAR RNA b
gb_ov:XLU88065	+ 224.00	226.89	0.0002	4975	! U88065 Xenopus laevis dsRNA ad
gb_ov:XLU88066	+ 221.00	226.23	0.0002	3773	! U88066 Xenopus laevis dsRNA ad
gb_ro:MMU79962	+ 219.50	232.95	8.1e-05	1520	! U79962 Mus musculus protamine-
gb_pr3:AF083033	+ 211.50	225.70	0.0002	1322	! AF083033 Homo sapiens PKR-assy
gb_pat1:AR028451	+ 211.50	223.91	0.0003	1608	! AR028451 Sequence 2 from paten
gb_pr3:AF072860	+ 211.00	222.65	0.0003	1740	! AF072860 Homo sapiens protein
gb_ro:AF083032	+ 208.50	220.85	0.0004	1584	! AF083032 Mus musculus PKR-assy

gb_ov:XLNRNABIPR	+ 207.00	220.66	0.0004	1358	! M96370 Xenopus laevis RNA bi
gb_pr7:HSRNATFET	+ 206.50	213.35	0.0010	2847	! X83973 H.sapiens mRNA for T
gb_pr6:HS26514	- 206.50	181.67	0.0579	90547	! Z84477 Human DNA sequence
gb_pr7:HSU10324	+ 202.50	207.18	0.0022	3505	! U10324 Human nuclear factor
gb_ro:RNU43534	+ 201.00	207.67	0.0021	2789	! U43534 Rattus norvegicus do

seq_name: gb_pr3:AF061939

seq_documentation_block:

LOCUS AF061939 3506 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,
complete cds.

ACCESSION AF061939
VERSION AF061939
KEYWORDS GI:4335948

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 3506)

AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and Desgroselliers,L.

TITLE Mammalian stauflen is a double-stranded RNA and tubulin binding

JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)

MEDLINE 99147057

REFERENCE 2 (bases 1 to 3506)

AUTHORS Desgroselliers,L., Wickham,L. and Luo,M.

TITLE Direct Submission

JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900

FEATURES Location/Montreal, Quebec H3T 1J4, Canada

source

gene

CDS

1. .3506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1. .3506
/gene="STAU"
409. .2142
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="stauflen protein"
/protein_id="AAD17531.1"
/db_xref="GI:4335949"
/translation="MSOVQVQVQNPBSAALSGSOLINKNQSLLSQPLMSIPSTSSLP
ENAGRPIONALPSASITSTSAASEITPTVELNALCMKLGKPMYKVPDPSRMOST
YNNMRGAYPPRYFPYPPVPLLYQVELSVGGQFNGKGTROAKHDAAKALRIL
ONEPLPERLEVNGRESEENLNKSEISQVEFIALKRLNPVNEVARESGPPhMKNEVT
KVSVEFVGESEKSKISKNAALVLEELKLPPLPAYEVKPRIKKTKPIYRQ
TSPEYGGGINSRLAQIOQAKKEPEYTLTERGLPRRREVMQVKGNTAEGRG
TNKKVAKRNAENMLEILGFKVPQPTKPAKSEKTPIKPGDGRKVTFFEPSGD
ENGTSNKEDEFRMPYLSHQQLPAGILPMVPEVAQAVGSQGHHTKDFTRAAPNPAKAT
VTAMIARELILYGGTSPYAEITLKNISSGHVPHGPLTRPSQDLYLSRVQGFQVEYKD
FPKNNKNEFVSLINCSQPPLISHGIGKDVESCHDMAALNITKLSELDOOSTEMPT
GNPMSVCGRC"

BASE COUNT 965 a 875 c 790 g 876 t
ORIGIN

alignment_scores:
Quality: 2970.00 Length: 577
Ratio: 5.147 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-316-048-6 x AF061939

Align seg 1/1 to: AF061939 from: 1 to: 3506

1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG1 17
|||||

409 ATGCTCAAGTTCAGTGCAGTTCAGAACCCATCTGCTCTCTCAGG 458
17 ySerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeuMets 34
|||||
459 GAGCCAATACTGTAACAAGAACCCAGTCTCTCTCACAGCCTTGATGA 508
34 erIleProSerThrThrSerSerLeuProSerGluAsnIleGlyArgPro 50
|||||
509 GTATTCTCTTACTACTAGCTCTCTGCCCTCTGAAATGCAGGTAGACCC 558
51 IleGlnAsnSerAlaLeuProSerAlaSerIleThrSerThrSerAlaAl 67
|||||
559 ATTCAAAACTCTGCTTTACCCCTCTGCATCTATTACATCCACCAGTGCAGC 608
67 aAlaGluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysL 84
|||||
609 TGCAGAAAGCATAAACCCCTACTGTAGAACTAATGCACCTGTGCTGAAC 658
84 euGlyLysLysProMetTyrLysProValAspProTyrSerArgMetGln 100
|||||
659 TTGAAAAAAACCAATGTATAGCCCTGTGACCCCTTACTCTCGGATGCAG 708
101 SerThrTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProArgTyrPh 117
|||||
709 TCCACCTATACTACACATGAGAGAGAGGCTTATCCCCCGAGGTACTT 758
117 eTyrProPheProValProProLeuLeuTyrGlnValGluLeuSerValG 134
|||||
759 TTACCCATTTCCAGTTCCACCTTACTTATCAAGTGGAACCTTCTGTGG 808
134 LyGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 150
|||||
809 GAGGACAGCAATTAAATGGCAAGAAAGAACAGACAGGCTGCGAAACAC 858
151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProG 167
|||||
859 GATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCCCCTGCCACA 908
167 uArgLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysS 184
|||||
909 GAGGCTGGAGGTGAATGGAAGAGAATCCGAAGAAATCTCAATTAAT 958
184 erGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProVal 200
|||||
959 CTGAATTAAGTCAAGTGTGAGATTGCACCTTAACGGAACCTTGCTGTG 1008
201 AsnPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheVa 217
|||||
1009 AATTTCGAGGTGGCCGGGAGAGTGGCCCAACCCACATGAAGAACTTTGT 1058
217 lThrLysValSerValGlyGluPheValGlyGluGlyLysSerL 234
|||||
1059 GACCAAGGTTTCGCTGGGGAGTTTGTGGGGGAAGGTGAAGGGAAGCA 1108
234 ySbLysIleSerLysLysAsnAlaAlaAlaValLeuGluGluLeuLys 250
|||||
1109 AGAAGATTTCAAAGAAAAATGCCGCCATAGCTGTCTTGAGGAGCTGAAG 1158
251 LysLeuProProLeuProAlaValGluArgValLysProArgIleLysLys 267
|||||
1159 AAGTTACCGCCCTGCTGCAGTTGAAGAGTAAAGCCTAGAAATCAAAAA 1208
267 sLysThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnG 284
|||||
1209 GAAAAACAACCACTAGTCAAGCCACAGACAAGCCCAAGAAATATGCCAGG 1258
284 LyIleAsnProIleSerArgLeuAlaGlnIleGlnAlaLysLysGlu 300
|||||
1259 GGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGCAAAAAAGAGAG 1308
301 LysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgAr 317
|||||
1309 AAGGAGCCAGAGTACAGCTCTCTCACAGAGCGAGGCTCCCGCGCGCAG 1358

317 gGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrG 334
|||||
1359 GGAGTTTGTGATGACAGGTGAAGGTTGGAACCAACACTGCAGAGAACGG 1408
334 lYThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 350
|||||
1409 GCACCAACAAGAGGTGGCCAAAGCCGCAATGCAGCCGAGAACATGCTGGAG 1458
351 lLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLys 367
|||||
1459 ATCCTTGTTCAAGTCCCGCAGCGGCAAGCCCAACCCGCACTCAA 1508
367 sSerGluGluLysThrProIleLysLysProGlyAspGlyArgLysValT 384
|||||
1509 GTCAGAGAGAGACACCCCATAAAGAAACCAAGGATGGAAGAAAGTAA 1558
384 hrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGlu 400
|||||
1559 CCTTTTGAACCTGGCTCTGGGATGAATAATGGAGTACTAGTAATAAGAG 1608
401 AspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyI 417
|||||
1609 GATGAGTTACAGATGCCCTTATCTAAGTCATCAGCAGCTGCCCTGCTGAAT 1658
417 eLeuProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyH 434
|||||
1659 TCTTCCATGTGTGCCGAGGTGCGCCAGCGCTGTAGAGATTAGTCAAGGAC 1708
434 lShiThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr 450
|||||
1709 ATCACACCAAAAGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACG 1758
451 ValThrAlaMetIleAlaArgGluLeuLeuTyrGlyLysThrSerProTh 467
|||||
1759 GTAACCTGCCATGATAGCCGAGAGTTGTGTATGGGGCACCCTGCCAC 1808
467 rAlaGluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisG 484
|||||
1809 AGCCGAGACCATTTAAAGAATAACATCTCTTACAGGCCACGTACCCCATG 1858
484 lYProLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGln 500
|||||
1859 GACCTCTCACGAGACCTCTGAGCAACTGGACTATCTTCCAGAGTCCAG 1908
501 GlyPheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPh 517
|||||
1909 GGATTCCAGGTTGAATACAAAGACTTCCCAAAAAACAACAAGAAAGAAATT 1958
517 eValSerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyI 534
|||||
1959 TGTATCTCTTATCAATTGCTCTCTCAGCCACCTGTGATCAGCATGTGTA 2008
534 lGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeu 550
|||||
2009 TCGCAAGAGATGAGAGTCCCTGCCATGATATGCGCTGCGCTGAACATCTTA 2058
551 LysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrG 567
|||||
2059 AAGTTGCTGTCTGAGTTGAGACCAACAAGTACAGAGATGCCAAGAACAG 2108
567 yAsnGlyProMetSerValCysGlyArgCys 577
|||||
2109 AAACGGACCAATGTCTGTGTGTGGGAGGTGC 2139

seq_name: gb_pr3:AF061940

seq_documentation_block:

LOCUS AF061940 3142 bp mRNA PRI 04-MAR-1999

DEFINITION Homo sapiens stauken protein (STAU) mRNA, alternatively spliced,
complete cds.

ACCESSION AF061940
VERSION AF061940.1 GI:4335950

```

KEYWORDS      '
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
TITLE          1 (bases 1 to 3142)
JOURNAL        Wickham, L., Duchaine, T., Luo, M., Nabl, I.R. and Desgroselliers, L.
AUTHORS        Mammalian staufen is a double-stranded RNA and tubulin binding
TITLE          protein which localizes to the rough endoplasmic reticulum
JOURNAL        Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE        99147057
REFERENCE      2 (bases 1 to 3142)
AUTHORS        Desgroselliers, L., Wickham, L. and Luo, M.
TITLE          Direct Submission
JOURNAL        Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
                Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES       Location/Qualifiers
                1. .3142
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="20"
                /map="20q13.1"
                1. .3142
                /gene="STAU"
                288. .1778
                /gene="STAU"
                /function="RNA binding protein"
                /note="alternatively spliced transcript"
                /codon_start=1
                /product="staufen protein"
                /protein_id="AADI7532.1"
                /db_xref="GI:4335951"
                /translation="MKLGKKPMKPYDPYSRMQSTYNNMRGAYPPRYFYPVPVPL
                LYQVELSVGGQOFNGKGTROAKHDAAKALRILQNEPLPERLEVNGRESEENLNK
                SEISOYFEIALKRNLPVNEFVARESGPPHMKNEVTKVSVEFVGEGSKSKISKNA
                AIAVLEELKKLPPLPAVERVKPRIKKTKPIVKPQTSPEYGGINPISRLAQIQAKK
                EKEPEYTLTERGLPRRREFVMQVGNHTAEGTGNKKVAKRNAENMLEILGEKVP
                QROPTPALKSEKTPRIKPGDGKRVTFEPGSGDENGTSNKKEDFRMPYLSHQQLPA
                GLPMVPEVAQAVGSQGHHTKDFTRAPNPAAKATVTAMIARELLYGGTSPTAETILK
                NNISGHHVPHGPLTRPSEQLDYLSRVQGFQVEYKDFPKNNKNEFVSLNCSQPPPLS
                HGIGKVESCHDMAALNIIKLSELDOOSTEMPRGTGNGPMSVCGRC"
BASE COUNT    862 a      778 c      733 g      769 t
ORIGIN
alignment_scores:
                Quality: 2664.50          Length: 563
                Ratio: 4.990              Gaps: 1
Percent Similarity: 94.849      Percent Identity: 93.073
alignment_block:
US-09-316-048-6 x AF061940 ..
Align seg 1/1 to: AF061940 from: 1 to: 3142
15 LeuSerGlySerGlnIleLeuAsnLysAsnGlnSerLeuSerGlnPr 31
   ||| ||| ::| ::| ::| ||||| ||
121 CTCACGCGCACCTCCGCTCTTCCCTCCCTTCGCTCCCTCTCC 170
   ||| ||| ::| ::| ::| ||||| ||
31 OleuMetSerIleProSerThrThrSerSerLeuProSerGluAsnAlaG 48
   |:::| ::| ||||| ::| ||| ||| ::::| |||
171 TTTTTCCTTCTTCTTCCCTCCCTTCCTCGCGCCACCGCCACGACGCG 219
   ||| |||
48 lYArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSerThr 64
   |||||
220 GCCGGGGG.....ACGAGTCCG 236
65 SerAlaAlaAlaGluSerIleThrProThrValGluLeuAsnAlaLeuCy 81
   ::| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 GAGCAGCAGCAGAAAGCATTAACCCCTACTGTAGAACTAAATGCACACTGTG 286
81 SmetLysLeuGlyLysLysProMetTyrLysProValAspProTyrSera 98
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

287	CATGAACCTTGAAAAAACC	CAATGTAT	AAGCCGTGTGACCCTTACTCTC	336
98	rgmEtGlnSerThr	TYrAsnTYrAsnMet	ArgGlyGlyAlaTYrProPro	114
337	GGATGCAGTCCACCCTAT	ACTACACATGAGAGAGAGGTGCTTATCCCCG		386
115	ArgTYrPheTYrProPhe	ProValProProLeuLeuTYrGlnValGluLe		131
387	AGGTACTTTTACC	CAATTCCAGTTCACCTTTACTTTATCAAGTGA	ACT	436
131	uSerValGlyGlyGlnGlnPhe	AsnGlyLYsGlyLYsThrArgGlnAla		148
437	TTCTGTGGAGACAGCA	CAATTAAATGGCAAGGAAAGACAAGACAGGCTG		486
148	lAlYsHisAspAlaAlaAla	lAlaLeuArglLeuGlnAsnGluPro		164
487	CGAACAACGATGCTGCTG	CCCAAGCGTTGAGGATCCTGCAGATGAGCCC		536
165	LeuProGluArgLeuGluVal	AsnGlyArgGluSerGluGluGluAsnLe		181
537	CTGCCAGAGAGCGCTGAG	GTGAATGGAAGAAATCCGAAGACAAATCT		586
181	uAsnLYsSerGluIleSerGln	ValPheGluIleAlaLeuLYsArgAsnL		198
587	CAATAAATCTGA	AATTAAGTCAAGTGTGAGATTGCACCTTAACGGA	ACT	636
198	euproValAsnPheGluVal	AlaAlaArgGluSerGlyProProHisMetLYs		214
637	TGCCTGTGAATTCGAGGTG	CGCCCGGAGAGAGTGGCCACCCACATGAAG		686
215	AsnPheValThrLYsValSer	ValGlyGluPheValGlyGluGlyGluG		231
687	AACTTTGTGACCAAGGTTT	CGTTGGGAGGTTTGTGGGGGAAGGTGAAGG		736
231	LYsSerLYsLYsIleSerLYs	LYsAsnAlaAlaIleAlaValLeuGluG		248
737	GAAAAACAAGAAGATTTC	ACAAGAAAAATGCCGCATAGCTGTTCTTGAGG		786
248	luleuLYsLYsLeuProPro	leuProAlaValGluArgValLYsProArg		264
787	AGCTGAAGAAGTTACCGCC	CTGCTGCAGTTGAACGAGTAAGCCTAGA		836
265	lIeLYsLYsLYsThrLYsPro	IleValLYsProGlnThrSerProGluTY		281
837	ATCAAAAAGAAAAACA	AAACCATAGTCAAGCCACAGACAAGCCAGAATA		886
281	rglYglnGlyIleAsnPro	IleSerArgLeuAlaGlnIleGlnAlaAl		298
887	TGGCCAGGGGATCAATCC	GAATTAGCCGACTGGCCAGATCCAGCAGGCAA		936
298	ysLYsGluLYsGluProGlu	TYrThrLeuLeuThrGluArgGlyLeuPro		314
937	AAAAGGAGAAGAGCCAG	ATACAGCTCTCACAGAGCGAGGCTCCCG		986
315	ArgArgArgGluPheVal	MetGlnValLYsValGlyAsnHisThrAlaG		331
987	CGCCGCAGGAGTTTGT	ATGATGCAGGTGAAGTTGGAACCACACTGCAGA		1036
331	uGlyTYrGlyTYrAsnLYs	LYsValAlaLYsArgAsnAlaIaGluAsnM		348
1037	AGGAACGGGACACCA	CAAGAGGTGGCCAAGCGCAATGCAGCCGAGAACA		1086
348	etLeuGluIleLeuGlyPhe	LYsValProGlnArgGlnProThrLYsPro		364
1087	TGCTGAGAGATCCTTG	TTCAAGTCCCGCAGCGGACGCCACCAAAACC		1136
365	AlaLeuLYsSerGluGlu	LYsThrProIleLYsLYsProGlyAspGlyAr		381
1137	GCACCTCAAGTCA	GAGAGACACCCATTAAGAAACCAGGGGATGGAAG		1186
381	gLYsValThrPhePheGlu	ProGlySerGlyAspGluAsnGlyTYrSera		398
1187	AAAAGTAACTTTT	TGAACTGGCTCTGGGGATGAATAATGGCACTAGTA		1236

398 snlysgluaspglupheargmetprotyrleuserhisglinglnleupro 414
1237 ATAAGAGAGATGAGTTCAGGATGCCCTTATCTAAGTCATCAGCAGCTGCCCT 1286
415 AlaglylleuPrometValProgluValAlaglnAlaValGlyValSe 431
1287 GCTGGAATTCCTCCCATGCTGCCGAGGTCGCCCAGGCTGTAGGAGTTAG 1336
431 rGlnGlyHisHisThrLysAspPheThrArgAlaAlaProasnProAla 448
1337 TCAAGGACATCACACCAGATTTTACCAAGGAGGAGCTCCGAATCCTGCCA 1386
448 ysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyThr 464
1387 AGGCACGCTAAGTCCATGATAGCCGAGAGTGTGTATGCGGCACACC 1436
465 SerProThrAlaGluThrIleLeuLysAsnIleSerSerGlyHisVa 481
1437 TCGCCACAGCCGAGACCATTTTAAGATAACATCTCTCAGGCCACGT 1486
481 lProHisGlyProLeuThrArgProSerGluGlnLeuAspTyrLeuSera 498
1487 ACCCCATGAGACCTCTCAGAGACCCCTCTGAGCAACTGACATCTTTCCA 1536
498 rGValGlnGlyPheGlnValGluTyrLysAspPheProLysAsnLys 514
1537 GAGTCCAGGAGTCCAGGTTGAATACAAAGACTTCCCCAAACACACAAG 1586
515 AsnGluPheValSerLeuIleAsnCySerSerGlnProProLeuIleSe 531
1587 AACGAATTTGTATCTCTTATCAATGTGCTCTCAGCCACCTCTGATCAG 1636
531 rHisGlyIleGlyLysAspValGluSerCySHisAspMetAlaAlaLeuA 548
1637 CCATGTATCGCAAGATGTGAGTCTGACATGATATGCTGCGCTGA 1686
548 snIleLeuLysLeuLeuSerGluLeuAspGlnSerThrGluMetPro 564
1687 ACATCTTAAGTTGCTGTCTGAGTTGGACCAACAAGTACAGATGCCA 1736
565 ArgThrGlyAsnGlyPrometSerValCysGlyArgCys 577
1737 AGAACAGGAACGACCAATGTCTGTGTGGAGGTGC 1775
seq_name: gb_pr3:AF061938
seq_documentation_block:
LOCUS AF061938 3217 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061938
VERSION AF061938.1 GI:4335946
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3217)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3217)
AUTHORS DesGroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1..3217
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="20"
/map="20q13.1"
1..3217
/gene="STAU"
363..1853
CDS
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AADI7530.1"
/db_xref="GI:4335947"
/translation="MKLGKKPMYKRPVDPYSRMOSTYNNMRGAYPPRYEYFPVPPL
LYOVELSVGGQFNGKTRQAAKHDAARILQNEPLPERLENGRESEENLNK
SEISOVFEIALKRNLPVNEVARESGPHMKNFVTKVSGEFVGEDEGSKISKNA
AIAVLEELKLPPLPAVERVKPRIKKTKPIYKPTSPYEGGGINPISRLAQIOQAK
EKEPEYTLTERGLPRRREFVMQVGNHTAEGTNNKVAKRNAENMLEILGFKV
OROPTKPAKSEEKTPIKKPGDGRKVTFFPESGDENGTNKEDEFRMPYLSHOQLPA
GILPMVPEVAQAVSQGHHTKDFTRAAPNPAKATVTAMIARELLYGTSPTAETILK
NNISSGHVPHGLTRPSEQLDYLRSVQGFQVEYKDFPKNNKNEFVSLINCSSOPPLIS
HGIGKDVESCHDMAALNILKLSLSDQSTEMPTGTNGPMVSGRC"
BASE COUNT 889 a 796 c 742 g 790 t
ORIGIN
alignment_scores:
Quality: 2659.00 Length: 577
Ratio: 4.906 Gaps: 3
Percent Similarity: 93.934 Percent Identity: 91.508
alignment_block:
US-09-316-048-6 x AF061938 ..
Align seg 1/1 to: AF061938 from: 1 to: 3217
15 leuserglysercInlleuAsnLysasnGlnSerleuSerGlnpr 31
121 CTCCACGCGCCACTCCGCTCTCCCTCCCTTCGTCCTCTCTCTCC 170
31 oleuMetSerIleProSerThrThrSerSerleuProserGluAsnAlag 48
171 TTTTTCCTTCTTCTCTCCCTCCCTCCCTCCCTCCCTCTCTCC 219
48 lYArg.....ProIleGlnAsn.....SerAlaLeuPro 57
220 GCCGGGGGAGCAGCTCCGGAGCAGCAGCAGAGTTATTAACTTAACC 269
58 SerAlaSerIleThrSerThrSerAla.....Al 67
270 TCTCAGAACTGACAAAGACAAACATGTTCTCTGGAACGCCCTTTTAA 319
67 aAlagIleSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysL 84
320 AAAAGAAAGCATAAACCCCTACTGTAGAACTAAATGCACCTGTCATGAAAC 369
84 euGlyLysLysPrometTyrLysProValAspProTyrSerArgMetGln 100
370 TTGGAAAAAACCAATGTATAGCCTGTGACCCCTTACTCTCGATGCAG 419
101 SerThrTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProArgTyrph 117
420 TCCACCTATACTACACATGAGAGAGGTGCTTATCCCGGAGGTACTT 469
117 eTyrProPheProValProProLeuLeuTyrGlnValGluLeuSerValG 134
470 TTACCCATTTCCAGTTCACACTTACTTATCAAGTGAACCTTCTGTGG 519
134 lYgIleGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 150
520 GAGGACAGCATTTAATGGCAAGGAACACACAGACAGGCTGCGAAACAC 569
151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGl 167

570 GATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCCTGCCAGA 619
167 uargleuGlValasnGlyargGluSerGluGluAsnLeuasnLysS 184
|||||
620 GAGGCTGGAGTGAATGGAAGAAATCCGAAGAGAAATCTCAATAAAT 669
184 erGlulSerGlnValPheGluIleAlaLeuLysArgAsnLeuProVal 200
|||||
670 CTGAATTAAGTCAAGTGTGAGATTGCACCTTAACGGAACCTTGCCCTGTG 719
201 AsnPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheVa 217
|||||
720 AATTTCGAGGTGGCCGGAGAGTGGCCACCCACATGAAGAACCTTGT 769
217 lThrLysValSerValGlyGluPheValGlyGluGlyGlyLysSerL 234
|||||
770 GACCAAGGTTTCGGTTGGGGAGTTTGTGGGGGAAAGGTGAAGGAAAGCA 819
234 yslLysIleSerLysLysAsnAlaAlaIleAlaValLeuGluLeuLys 250
|||||
820 AGAAGATTTCAAGAAATGCCGCCATAGCTGTCTTGAGGAGCTGAAG 869
251 LysLeuProProLeuProAlaValGluArgValLysProArgIleLysLy 267
|||||
870 AAGTTACCGCCCTGCTGCAGTTGAACGAGTAAGCCTAGAAATCAAAAA 919
267 sLysThrLysProIleValLysProGlnThrSerProGluTyrGlyGln 284
|||||
920 GAAACAAACCAACCATAGTCAAGCCACAGACACAGCCCAAGAATATGCCAG 969
284 lYleAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlu 300
|||||
970 GGATCAATCCGATTAGCCGACTGGCCAGATCCAGACGCAAAAAAGAGAG 1019
301 LysGluProGluTyrThrLeuLeuThrgLuarGlyLeuProArgArgAr 317
|||||
1020 AAGGAGCCAGAGTACAGCTCCTCACAGAGCGAGCGCTCCCGCGCCAG 1069
317 gGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrg 334
|||||
1070 GGAGTTTGTGATGACAGTGAAGGTTGGAAACACACTGCAGAGAAAGCGG 1119
334 lYThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 350
|||||
1120 GCACCAACAAGAAGGTGGCCAAAGCGCAATGCAGCCGAGAATGCTGGAG 1169
351 IleLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLy 367
|||||
1170 ATCCTTGCTTCAAGTCCCGCAGCGGACGCCCAACCCGCACTCAA 1219
367 sSerGluGluLysThrProIleLysLysProGlyAspGlyArgLysValT 384
|||||
1220 GTCAGAGGAGAAAGACACCATTAAGAAACCAAGGATGGAAGAAAAGTAA 1269
384 hrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGlu 400
|||||
1270 CCTTTTGAACCTGCTCTGGGATGAAAAATGGACTAGTAATAAAGAG 1319
401 AspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyI 417
|||||
1320 GATGAGTTTCAGGATGCCCTATCTAAGTCATCAGCAGCTGCTGGAAT 1369
417 eleuProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyH 434
|||||
1370 TCTTCCCATGTGTCGCCGAGGTGCCAGGCTGTAGAGATTAGTCAAGGAC 1419
434 iShIsthrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr 450
|||||
1420 ATCACAACAAGATTTTACCAGGGCAGCTCCGAATCTGCCAAGGCCACG 1469
451 ValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProTh 467
|||||
1470 GTAACTGCCATGATAGCCCGAGAGTTGTTGTATGGGGCACCTCGCCAC 1519

467 ralagluThrIleLeuLysAsnIleSerSerGlyHisValProHisG 484
|||||
1520 AGCCGACACCATTTTAAGATAACATCTCTTCAGGCCACGTACCCCATG 1569
484 lYProLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGln 500
|||||
1570 GACCTCTCAGAGACCCCTTGAGCAACTGACTATCTTTCCAGAGTCCAG 1619
501 GlyPheGlnValGluTyrLysAspPheProLysAsnLysAsnGluPh 517
|||||
1620 GGATTCAGGTTGAATACAAAGACTTCCCAAAAAACAAGAAGCAAT 1669
517 eValSerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyI 534
|||||
1670 TGTATCTCTTATCAATTCCTCTCTCAGCCACCTGTGATCAGCCATGTA 1719
534 legLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeu 550
|||||
1720 TCGGCAAGATGTGAGAGTCTCCATGATATGGCTGCGCTGAACATCTTA 1769
551 LysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGl 567
|||||
1770 AAGTTCTGCTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGAACAG 1819
567 yAsnGlyProMetSerValCysGlyArgCys 577
|||||
1820 AAACGACCAATGTCTGTGTGGAGGTGC 1850
seq_name: gb_pr3:AF061941
seq_documentation_block: 3349 bp mRNA PRI 04-MAR-1999
LOCUS AF061941
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061941
VERSION AF061941.1 GI:4335952
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3349)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroseillers,L.
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3349)
AUTHORS DesGroseillers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source location/Qualifiers
1..3349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1..3349
/gene="STAU"
495..1985
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AADI7533.1"
/db_xref="GI:4335953"
/translation="MKLGKKPMYKPYDYSRMOSTYNNYMGAYPPRYFPVPVPL
LYQVELSVGGQFNGKGTROAAKHDAAKALRIQNEPLERLELVNGRESEENLNK
SEISQVEFIALKRNLPVNFVARRESGPHMKNFVTKVSVEFVGEVGEKSKISKNA
AIVALEELKKLPPLPAVERVKPRIKRTKPIVKPQTSPEYGGGINPISRLAQIQAKK

EKEPEYTLTERGLPRRREFVMQVGNHTAEGTGNKKVAKRNAENMLEILGFKVP
OROPTKPAKSEKTPDKKPGDGRKVTFFEPGSGDENGTSNKEDEFMPYLSHOOLPA
GILPMVPEVAQAVGVSQGHHTKDFTRAAPNPAKATVTAMIARELLYGSTPTAETILK
NMISSGHVPHGLTRPSEQLDYLSRVQGFQVEYKDFPKNNKNEFVSLINCSSQPLIS
HGIGDVESCHDMAALNLIKLSLSDQSTEMPTGTNGPMSVCGRC"
BASE COUNT 913 a 838 c 777 g 821 t
ORIGIN

alignment_scores:
Quality: 2649.00 Length: 509
Ratio: 5.204 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-316-048-6 x AF061941 ..

Align seg 1/1 to: AF061941 from: 1 to: 3349

69 GluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysLeuG 85
|||||
456 GAAAGCATAAACCCCTACTGTAGAACTAAATGCACCTGTCATGAACCTGG 505
85 YLysLysProMetTyrLysProValAspProTyrSerArgMetGlnSer 102
|||||
506 AAAAAAACCAATGTATAGCCTGTGACCCCTTACTCTCGGATGCAGTCCA 555
102 hTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProArgTyrPheTyr 118
|||||
556 CCTATACTACACATGAGAGAGAGGTGCTTATCCCGAGGTACTTTTAC 605
119 ProPheProValProProLeuLeuTyrGlnValGluLeuSerValGlyG 135
|||||
606 CCATTTCCAGTTCACACCTTTACTTATCAAGTGAAGTCTTCTGCGGAGG 655
135 yGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAsp 152
|||||
656 ACAGCAATTTAATGGCAAAGGAAGACACAGACAGCTGCGAAACAGCATG 705
152 IaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArg 168
|||||
706 CTGCTGCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCCTGCCAGAGAGG 755
169 LeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSerG 185
|||||
756 CTGAGAGTGAAATGGAAGAGAATCCGAAGAGAAATCTCAATAAATCTGA 805
185 uIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsn 202
|||||
806 AATAAGTCAAGTGTGTGAGATTGCACCTTAAACGGAACCTTGCCCTGAATT 855
202 heGluValAlaArgGluSerGlyProProHisMetLysAsnPheValThr 218
|||||
856 TCGAGGTGGCCGGGAGAGTGGCCACCCACATGAAGAACTTTGTGACC 905
219 LysValSerValGlyGluPheValGlyGluGlyLysSerLysLy 235
|||||
906 AAGGTTTCGGTGGGAGAGTTGTGGGGGAAGGTGAAGGGAAGCAAGAA 955
235 sIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLysLysL 252
|||||
956 GATTTCAAAGAAAAATGCCGCATAGCTGTCTTGAGGAGCTGAAGAACT 1005
252 euProProLeuProAlaValGluArgValLysProArgIleLysLysLys 268
|||||
1006 TACCGCCCTGCTGCAAGTTGAACGAGTAAGCCTAGAAATCAAAAGAAA 1055
269 ThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnGlyI 285
|||||
1056 ACAAAACCATAGTCAAGCCACAGACAGCCCAAGATATGCGCAGGGGAT 1105
285 eAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysG 302
|||||

1106 CAATCCGATTAGCCGACTGGCCACAGATCCAGACAGGCAAAAAAGAGAAAG 1155
302 LuProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgGlu 318
|||||
1156 AGCCAGAGTACACGCTCTCACAGAGGAGGCTCCCGCGCCGAGGGAG 1205
319 pheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyTh 335
|||||
1206 TTTGTGATGACAGTGAAGGTTGGAAAACCACTGCAGAGGAACGGGCAC 1255
335 rAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIle 352
|||||
1256 CAACAAGAAAGTGGCCAAAGCCCAATGCAGCCGAGAAACATGCTGAGATCC 1305
352 euGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLysSer 368
|||||
1306 TTGGTTTCAAGTCCCGCAGCGGCGAGCCCAACCAACCCGACTCAAGTCA 1355
369 GluGluLysThrProIleLysLysProGlyAspGlyArgLysValThrPh 385
|||||
1356 GAGGAGAAGACACCCATTAAGAAACCAGGGGATGGAAGAAAAGTAACTTT 1405
385 epheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspG 402
|||||
1406 TTTTGAACCTGGCTCTGGGATGAAAATGGGAACTAGTAATAAAGAGATG 1455
402 LuPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeu 418
|||||
1456 AGTTCAGGATGCCCTTATCTAAGTATCAGCAGCTGCCTGCTGAATTCTT 1505
419 ProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyHisH 435
|||||
1506 CCCATGGTGCCCGAGGTGCGCCAGGCTGTAGAGATTAGTCAAGGACATCA 1555
435 sThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrVal 452
|||||
1556 CACCAAAAGATTTAACGAGGCGAGCTCCGAATCTGCCAAAGGCCACGGTAA 1605
452 hrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAla 468
|||||
1606 CTGCCATGATAGCCGAGAGTGTGTGTATGGGGGCACCTCGCCACAGCC 1655
469 GluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisGlyPr 485
|||||
1656 GAGACCATTTTAAGAATAACATCTCTTCAGGCCACGTACCCCATGGACC 1705
485 OLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGlnGlyP 502
|||||
1706 TCTCAGGAGACCTCTGAGCAACTGGACTATCTTTCAGAGATCCAGGGAT 1755
502 heGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheVal 518
|||||
1756 TCCAGGTTGAATACAAAGACTTCCCAAAAAACAACAAGAACGAATTTGTA 1805
519 SerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyIleGl 535
|||||
1806 TCTCTTATCAATTGCTCTCTCAGCCACCTCTGATCAGCCATGTATCGG 1855
535 YLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysL 552
|||||
1856 CAAGGATGTGAGTCTCCATGATATGCTGCGCTGAACATCTTAAAGT 1905
552 euLeuSerGluLeuAspGlnGlnInsThrGluMetProArgThrGlyAsn 568
|||||
1906 TGCTGTCTGAGTTGACCAACAACAAGTACAGAGATGCCAAGAACAAGAAAC 1955
569 GlyProMetSerValCysGlyArgCys 577
|||||
1956 GGACCAATGTCTGTGTGTGGAGGTGC 1982
seq_name: gb_pr4:AK025519
seq_documentation_block:

LOCUS AK025519 3194 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21866 fis, clone HEP02379, highly similar to
AF061938 Homo sapiens staufen protein (STAU) mRNA.
ACCESSION AK025519
VERSION AK025519.1 GI:10438059
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 CDNA to mRNA, clone_lib:HEP
clone:HEP02379.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE MEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 3194)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT MEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Departent of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source 1. Location/Qualifiers
1.3194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP02379"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
misc_feature 1.3194
/note="highly similar to AF061938 Homo sapiens staufen
protein (STAU) mRNA"
BASE COUNT 894 a 787 c 732 g 781 t
ORIGIN
alignment_scores:
Quality: 2587.50 Length: 602
Ratio: 4.739 Gaps: 4
Percent Similarity: 90.698 Percent Identity: 87.375
alignment_block:
US-09-316-048-6 x AK025519 ..
Align seg 1/1 to: AK025519 from: 1 to: 3194
11 ProSerAlaAlaLeuSer..... 16
111:::||||| 111
15 CCGCGGCTGCGCGCTCTCTCGGCTCCCGCTTCCTTGACCGCCTCCC 64
17GlySerGlnIleLeuAsnLysAsn..... 24
111:::||||| 111
65 CCGCGGCGCGCGCGCGCGCTCTCTCCACGCGCCACTCCGCTCTTCCC 114
25GlnSerLeuLeuSerGlnProLeuMetSerIleProSerThrThr 39
:::||||| 111:::|
115 TCCCTTCGTCCCTTCTCTCCCTTTTTCCTTCTCTTCCCTTCCCC.TCC 163
40 SerSerLeuProSerGlnAsnAlaGlyArgProIleGlnAsnSerAlaLe 56

111 ||| :::::||||| :::::|
164 TCGCGCGCACCGCCAGACGACCGCGGGGAGACGCTCGAGCAGCA 213
56 uPro.....SerAlaSerIleThrSerThrSera 66
111 ::| :::::|
214 GCCAGAGTTTATTAACTTAACTCTCAGAACTGACAAAGACACAT 263
66 la.....AlaAlaGlnSerIleThrProThrVal 75
264 TGTTCTGGACGCGCTCTTTTAAAAAGAAAGCATTAACCTACTGTA 313
76 GluLeuAsnAlaLeuCysMetLysLeuGlyLysLysProMetTyrLysPr 92
|||||
314 GAACATAATGCACGTGCATGAACCTTGAAAAAACAATGATTAAGCC 363
92 oValAspProTyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArg 109
|||||
364 TGTTGACCTTACTCTCGATGCAGTCCACCTATATACATCAACATGAG 413
109 LylGlyAlaTyrProProArgTyrPheTyrProPheProValProProLeu 125
|||||
414 GAGGTGCTTATCCCGGAGGTACTTTTACCATTTCAGTTCACCTTTA 463
126 LeuTyrGlnValGluLeuSerValGlyGlnGlnPheAsnGlyLysG 142
|||||
464 CTTTATCAAGTGAACCTTCTGTGGAGAGACAGCAATTTAATGCCAAAG 513
142 YLstYrArGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuArg 159
|||||
514 AAAGACAGACAGCGCTGCGAAACGATGCTGTCGCAAAAGCTTGAGGA 563
159 leLeuGlnAsnGluProLeuProGluArgLeuGluValAsnGlyArgL 175
|||||
564 TCCTGCAGATGAGCCCTGCGAAGAGGCTGAGGTGAATGGAAGAA 613
176 SerGluGlnGluAsnLeuAsnLysSerGluIleSerGlnValPheGlu 192
|||||
614 TCCGAAGAGAAAAATCTCAATAAATCTGAATAATAGTCAAGTGTGAGAT 663
192 eAlaLeuLysArgAsnLeuProValAsnPheGluValAlaAlaArgGlu 209
|||||
664 TGCACCTTAACGGAACCTTGCCCTGTAATTTCGAGGTGCGCGGAGAGTG 713
209 lYProPheHisMetLysAsnPheValThrLysValSerValGlyGluP 225
|||||
714 GCCCACCACATGAAGAACTTTGTGACCAAGGTTTCGTTGGGGAGTTT 763
226 ValGlyGlnGlyGluGlyLysSerLysLysIleSerLysLysAsnAla 242
|||||
764 GTGGGGAGGTGAAGGGAAGCAAGAAATTTCAAAAGAAAAATGCCGC 813
242 AlLeAlaValLeuGluGluLeuLysLysLeuProProLeuProAlaVal 259
|||||
814 CATAGCTGTCTTGAGAGCTGAAGAGTTACCGCCCTGCTGCAGTTG 863
259 lUArgValLysProArgIleLysLysLysThrLysProIleValLysPro 275
|||||
864 AACGAGTAAGCCTAGAAATCAAAAAGAAACAAACCAATA.GTCAAGCCA 912
276 GlnThrSerProGluTyrGlyGlnGlyIleAsnProIleSerArgLeuAl 292
|||||
913 CAGACAAGCCAGAAATATGGCCAGGGGATCAATCCGATTAGCCGACTGGC 962
292 aglIleGlnGlnAlaLysLysGluLysGluProGluTyrThrLeuLeu 309
|||||
963 CCAGATCCAGCAGCAAAAAAGAGAGAGAGCAGAGTACAGCTCCTCA 1012
309 hrGluArgGlyLeuProArgArgArgGluPheValMetGlnValLysVal 325
|||||
1013 CAGAGCAGGCGCTCCCGCGCGCAGGAGTTTGTGATGACGAGGTGAAGGTT 1062
326 GLYAsnHisThrAlaGluGlyThrGlyThrAsnLysLysValAlaLysAr 342
|||||

1063 GGAACACCACTGCAGAGGAGGACGGGACCAACAAGAGGTGGCCAGCG 1112
342 gAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysValProGlnA 359
1113 CAATGCAGCCGAGAACATCTGGAGATCCTTGCTTCAAAAGTCCCGCAGG 1162
359 rgGlnProThrLysProAlaLeuLysSerGluGluLysThrProIleLys 375
1163 CGCAGCCCAACCAACCCGCACTCAAGTCAGAGAGAAGACACCCATAAAG 1212
376 LysProGlyAspGlyArgLysValThrPhePheGluProGlySerGlyAs 392
1213 AAACCAAGG.GATGGAAGAAAGTAACCTTTTGAACCTGGCTCTGGGGA 1261
392 pGluAsnGlyThrSerAsnLysGluAspGluPheArgMetProTyrLeus 409
1262 TGAATAATGGAGCTAGTAATAAAGAGATGAGTTCAGGATGCCCTTATCTAA 1311
409 eRhISGlnGlnLeuProAlaGlyIleLeuProMetValProGluValAla 425
1312 GTCATCAGCAGCTGCCCTGCTGGAATTCCTCCCATGCTGCCCGAGGTGCC 1361
426 GlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArgAl 442
1362 CAGGCTGTAGAGTAACTCAAGACATCACACCAAGATTTTACCAGGCG 1411
442 alaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluL 459
1412 AGCTCCGAATCCTGCCAAGGCCACCGTAACGCCATGATAGCCCGAGAGT 1461
459 euleuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsn 475
1462 TGTGTATGGGGGACCTCGCCACAGCCGAGACCATTTTAAAGATAAAC 1511
476 IleSerSerGlyHisValProHisGlyProLeuThrArgProSerGluGl 492
1512 ACCTCTTCAGGCCACGTAACCCCATGGACCTCTCACGAGACCCCTTGAGCA 1561
492 nLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAspP 509
1562 ACTGACTATCTTCCAGAGTCCAGGATCCAGGTTGAATACAAAGACC 1611
509 heProLysAsnAsnLysAsnGluPheValSerLeuIleAsnCySerSer 525
1612 TCCCCAAACCAACAAGACGAATTTGTAATCTTATCAATGTCTCTCT 1661
526 GlnProProLeuIleSerHisGlyIleGlyLysAspValGluSerCysHi 542
1662 CAGCCACCTCTGATCAGCCATGGTATCGGCAAGATGTGAGTCTGCCA 1711
542 sAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLeuAspGlnG 559
1712 TGATATGGCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGACCAAC 1761
559 InSerThrGluMetProArgTyrThrGlyAsnGlyPrometSerValCysGly 575
1762 AAAGTACAGAGATGCCAAGAACAGGAACGACCAATGTCTGTGTGGG 1811
576 ArgCys 577
1812 AGGTGC 1817

seq_name: gb_pf6: HSA132258

seq_documentation_block: 3066 bp mRNA PRI 14-MAY-1999
LOCUS HSA132258
DEFINITION Homo sapiens mRNA for stauflen protein, partial.
ACCESSION AJ132258
VERSION AJ132258.1 GI:4572587
KEYWORDS stauflen gene; stauflen protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 3066)
AUTHORS Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
TITLE Interaction of influenza virus NS1 protein and the human homologue of stauflen in vivo and in vitro
JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE 99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS Ortin,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology, Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN
FEATURES
source 1. .3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
36. .1526
/gene="stauflen"
36. .1526
/gene="stauflen"
/function="dsRNA binding protein"
/codon_start=1
/product="stauflen protein"
/protein_id="CAB40082.1"
/db_xref="GI:4572588"
/translation="MKLGKPMYKPPVDPYSRMOSTYNYNMRGAYPPRYEYPPVPPL
LYVELSVGGQGFNGKGTROAKHDAAKALRLIONEPLPERLEVNGRESEENLNK
SEISQVEIALKRNLPVNEVARESGPHMKNFVTKVSGFVGEGRSKISKNA
AIVALELKKLPPLPAVERVKPRIKKTTPYKQTSPEYGGINPIISRLAQIQAKK
EKEBEYTLTERGLPRREFVMQVGNHTAEGTGNKKVAKRNAENMLEILGFKP
QRSHQRTQVRGEDTHKTRGWKKSNFLNLALGMENGTSNKEDFRMPYLSHQQLP
GILPMVEVAQAVGVSOGHHTKDFTRAAPNPAKATVTAMIARELYGTSPTAETILK
NNISSGHVPHGPLTRPSEQLDILSRVQGFQVEYKDFPKNNKNEFVSLINCSQPLIS
HGIGKDVESCHDMAALNLIKLSLSDQSTEMPRTGNPMSVCGRC"
BASE COUNT 891 a 693 c 699 g 783 t
ORIGIN
alignment_scores:
Quality: 2577.50 Length: 510
Ratio: 5.104 Gaps: 4
Percent Similarity: 99.020 Percent Identity: 99.020
alignment_block:
US-09-316-048-6 x HSA132258 ..
Align seg 1/1 to: HSA132258 from: 1 to: 3066
71 IleThrProThrValGluLeuAsnAlaLeuCysMetLysLeuGlyLysLy 87
3 ATAACCCCTACTGTAGAACTAAATGCACCTGTGCATGAAACTTGGAAAAA 52
87 sPrometTyrLysProValAspProTyrSerArgMetGlnSerThrTyrA 104
53 ACCAATGTATAAGCCTGTGACCCCTTACTCTGGATGCAGTCCACCTATA 102
104 snTyrAsnMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhe 120
103 ACTACACATGAGAGAGGTGCTTATCCCGAGGTACTTTTACCACATT 152
121 ProValProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGln 137
153 CCAGTTCACACCTTACTTTATCAAGTGAACCTTCTGTGGAGAGACAGCA 202
137 npHeAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspAlaAla 154
203 ATTTAATGGCAAGGAAAGACACAAGACAGGCTGCCAAGACAGATGCTGCTG 252
154 lAlysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArgLeuGlu 170
253 CCAAGCGTTGAGATCCTGCAGAAATGAGCCCTGCGCAGAGAGAGCTGAG 302
171 ValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSerGluIleSe 187

|||||
303 GTGATGGAAGAAATCCGAGAGAAGAAAATCTCAATAAATCTGAATTAAG 352
187 rGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluV 204
|||||
353 TCAAGTGTGTGAGATGCACTTAACGGAACCTGCCTGTGAATTCGAGG 402
204 aIlaAArgGluSerGlyProProHisMetLysAsnPheValThrLysVal 220
|||||
403 TGGCCCCGGAGAGTGGCCCCACATGAAGAACTTTGTGACCAAGGTT 452
221 SerValGlyGluPheValGlyGluGlyGlyLysSerLysLysIleSe 237
|||||
453 TCGGTTGGGAGTTTGTGGGGAGGTGAAGGGGAAAGCAAGAGATTTC 502
237 rLysLysAsnAlaAlaIleAlaValLeuGluLeuLysLysLeuProp 254
|||||
503 AAAGAAAATGCCCGCATAGCTGTCTTGAGAGAGCTGAAGAAGTTACCGC 552
254 rOleuProAlaValAlaGluArgValLysProArgIleLysLysThrLys 270
|||||
553 CCTGCCTGCAGTTGAACGAGTAAGCCTAGAAATCAAAAAGAAAACAAA 602
271 ProIleValLysProGlnThrSerProGluTyrglyGlnGlyIleAsnPr 287
|||||
603 CCCATAGTCAAGCCACAGACAGCCAGAAATATGGCCAGGGGATCAATCC 652
287 oIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProG 304
|||||
653 GATTAGCCGACTGGCCCCAGATCCAGCAGGCAAAAAGAGAAGAGCCAG 702
304 lUTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgGluPheVal 320
|||||
703 AGTACAGCTCCTCACAGAGGAGGCTCCCGCGCCAGGAGGAGTTGTG 752
321 MetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnLy 337
|||||
753 ATGCAGGTGAAGTTGGAACCACTGACAGAAGGAGGGCACCAACAA 802
337 sLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyP 354
|||||
803 GAAGTGGCCCAAGCGCAATGCAGCCGAGACATGCTGAGATCCTTGCTT 852
354 heLysValProGlnArgGlnPro.ThrLysProAlaLeuLysSerGluG 370
|||||
853 TCAAGTCCCGCAGCGC..AGCCACCAAAACCGCCTCAAGTCAGAGGA 899
370 uLysThrProIleLysLysProGlyAspGlyArgLysValThr.PhePhe 386
|||||
900 GAAGACACCCATAAAGAACCAGGGGATGGAAGAAAGTAACCTTTT 949
387 GluProGlySerGlyAsp.GluAsnGlyThrSerAsnLysGluAspGluP 403
|||||
950 GAACCTGGCTCTGGGGATGGAATAATGGACTAGTAATTAAGAGATGAGT 999
403 heArgMetProTyrrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPro 419
|||||
1000 TCAGGATGCCCTTATCTAAGTCATCAGCAGCTGCTGCTGGAATTCTTCCC 1049
420 MetValProGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisHis 436
|||||
1050 ATGTGCCCGAGGTGCGCCAGGCTGTAGAGTTAGTCAAGACATCACAC 1099
436 rLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThra 453
|||||
1100 CAAAGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGCCACGCTAACTG 1149
453 lAmetIleAlaArgGluLeuLeuTyrglyGlyThrSerProThrAlaGlu 469
|||||
1150 CCATGATAGCCCGAGAGTGTGTATGGGGCACCTCGCCACACAGCCGAG 1199
470 ThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisGlyProLe 486
|||||

1200 ACCATTTTAAGAATAACATCTCTTCAGGCCACGATACCCCATGGACCTCT 1249
486 uThrArgProSerGluGlnLeuAspTyrrLeuSerArgValGlnGlyPheG 503
|||||
1250 CACGAGACCTCTGAGCAACTGCACTATCTTTCCAGAGTCCAGGATTC 1299
503 lNValGluTyrrLysAspPheProLysAsnLysAsnGluPheValSer 519
|||||
1300 AGTTGATATACAAAGACTTCCCAAAAACAACAAGACGAATTTGTATCT 1349
520 LeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyIleGlyLy 536
|||||
1350 CTTATCAATTGCTCTCTCAGCCACCTCTGATCAGCCATGTATCGGCAA 1399
536 sAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeu 553
|||||
1400 GGATGTGAGTCTCTGCCATGATATGGCTGCGCTGAACATCTTAAAGTTGC 1449
553 euSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGly 569
|||||
1450 TGTCTGAGTTGGACCAACAAGTACAGAGATGCCAAGACAGGAAACGGA 1499
570 PrometSerValCysGlyArgCys 577
|||||
1500 CCAATGTCTGTGTGTGGAGGTGC 1523
seq_name: gb_ro:AF290989
seq_documentation_block: 3041 bp mRNA ROD 09-AUG-2000
LOCUS AF290989 3041 bp mRNA
DEFINITION Rattus norvegicus staufen isoform Stau-I6 (Stau) mRNA, complete
ACCESSION AF290989
VERSION AF290989.1 GI:9755162
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3041)
AUTHORS Monshausen,M., Putz,U., Rehbein,M., Schweizer,M.,
Desgroesellers,L., Kuhl,D., Richter,D. and Kindler,S.
TITLE Two rat brain staufen isoforms differentially bind RNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3041)
AUTHORS Monshausen,M. and Kindler,S.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) Institute for Cell Biochemistry and
Clinical Neurobiology, University of Hamburg, University Hospital
Eppendorf, Suderfeldstr. 24, Hamburg D-22529, Germany
FEATURES
source location/Qualifiers
1..3041
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1..3041
/gene="Stau"
227..1696
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila staufen;
contains four RNA-binding domains; rStau-I6"
/codon_start=1
/product="staufen isoform Stau-I6"
/protein_id="AAF98177.1"
/db_xref="GI:9755163"
/translation="MYKPVDPHSRMQSTYSYGMRGAYPPRYFPFPVPLLYOVELS
VGQGFNGKGMRPVKHDATARALRTIOSEPLPERLEVNGRESEENLNKSEISOVF
EIALKRNLPVNFEVARESGPRHMKNFYRVSVGEFVGEGEKSKISKNAARAVLEQ
LRLRPLPAVERVKPRIKKKSQPTCKLOTAPDYGGMNPISRLAIOQAKKEPEYV
LITERGLPRRREVMQVKVGHHTAEGAGTNKKVAKRNAEENMLEILGKVPQAQPAKP

1548 TCGTCAGCCATGGCATCGGCAAGATGTGAGTCCCTGTTCATGATATGGCT 1597
546 AlaleuAsnIleLeuLysLeuSerGluLeuAspGlnGlnSerThrG1 562
1598 GCACTGACATCTTAAGCTGCTGTCTGAGTTGGACCAACAGACACAGA 1647
562 uMetProArgThrGlyAsnGlyProMetSerValCysGlyArgCys 577
1648 GATGCCAAGACAGGAACGACGAGTTTCAGCGTGGGAGGTGC 1693

seq_name: gb_ro:AF227200

seq_documentation_block:

LOCUS AF227200 3059 bp mRNA ROD 09-AUG-2000
DEFINITION Rattus norvegicus staufen isoform Stau+I6 (Stau) mRNA, complete cds, alternatively spliced.

ACCESSION AF227200 GI:9754865

VERSION AF227200.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

FEATURES

source

1. .3059

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/tissue_type="brain"

/dev_stage="adult"

1. .3059

/gene="Stau"

227. .1714

/gene="Stau"

/note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau+I6"

/codon_start=1

/product="staufen isoform Stau+I6"

/protein_id="AAF98119.1"

/db_xref="GI:9754865"

/translation="MYKVPDPSRMOSTYSYGMRGAYPPRYFYPPPPVPLLYOVELS VGGQFNGKGMKRPVKKHDATARALRTIOSEPLPERLEVNGRESEENLNKSEISQVF EIALKRNLVNFESFPLTOVARESPPHMKNFVTRVSVGEFVGESEKSKISKNNAA RAVLEQLRLPLPAVERVKPRIKKKSQPTCKLQTAPDYGGMNPISRLAQIQAKKE KEPEYMLTERGLPRRREFVMQVKGHTAGAGTNNKQVAKRNAEMLEILGFKVPQ AQPAPALKSEKTPVKKPGDGKRVTFEFPSPGDENGTNKEDEFRMPLSHQQLPAG ILPMPEVAQAVGVSQGHHTKDFARAAPNPAKATVTAMIARELLGGTSPTAETILKS NISSGHVPHGPRTRPSEQLYLSRAQGFQVEYKDFPKNNKNECVSLINCSOSPPLVSH GIGKDVESCHDMAALNILKLLSELDOQSTEMPRRTGNGPVSACGRC"

polyA_signal

3043. .3048

/gene="Stau"

BASE COUNT 817 a 778 c 749 g 713 t 2 others
ORIGIN

alignment_scores:

Quality: 2432.50 Length: 572
Ratio: 4.687 Gaps: 5
Percent Similarity: 90.734 Percent Identity: 83.566

alignment_block:

US-09-316-048-6 x AF227200 ..

Align seg 1/1 to: AF227200 from: 1 to: 3059

31 ProLeuMetSerIlePro.....SerThrTh 39
7 CCCCCTTCTTGACCCCGCGCGCGCGCGCGCTTCTCCGACGCCAC 56
39 rSerSerLeuProSerGluAsn..... 46
57 TCCGTCCTTCCCTCCCGCGCTTCTTCTCCCTCGCCCTTCTTGTGCT 106
47 ..AlaGlyArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThr 62
107 TTCTCTCCCTCGCGCGCGCGCAATACCGCGCGCGCGG.....ACG 147
63 SerThrSerAlaAlaAla.GluSerIleThrProThrValGluLeuAsn 79
148 AGCTCTGGGGAAGCAGCCAGAAAGTATAGCTTCTACCATTTGAGCTCAATG 197
79 lAlaCysMetLysLeuGlyLysLysProMetTyrLysProValAspPro 95
198 CACTGTGCGTGAACCTGGAAGAAACCAATGATATAAGCCCGTGAACCC 247
96 TyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArgGlyAlaTyr 112
248 CACTCTCGGATGAGTCCACCTACAGCATATGCGATGCGTGAGTGCCCTA 297
112 rProProArgTyrPheTyrProPheProValProProLeuLeuTyrGln 129
298 TCCCCCAGATACTTTTACCCATTCCACGTTCCACCGTTACTTACCAAG 347
129 aGluLeuSerValGlyGlyGlnGlnPheAsnGlyLysGlyThrArg 145
348 TTGAGCTCTCCGTGGGAGAGACAGACAGTTTAATGGGAAGGGAAGATGAGA 397
146 GlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAs 162
398 CCACCTGTGAACATGATGCCACTGCCGAGCCCTGAGAGACTGTGCAGAG 447
162 nGluProLeuProGluArgLeuGluValAsnGlyArgGluSerGluGlu 179
448 TGAGCCCTTACGAGAGAGAGCTAGAGGTAATGGAAGAGATCAGAGAAG 497
179 lAsnLeuAsnLysSerGluIleSerGlnValPheGluIleAlaLeuLys 195
498 AAAACCTTAATTAATCAGAATAAGCCAGTGTGAGATTGCCGTGAAG 547
196 ArgAsnLeuProValAsnPheGlu.....ValAlaAr 206
548 CGGAATTTGCTGTGAATTTTGAGTCTTTCCTCTGACACAGAGTGGCCG 597
206 gGluSerGlyProProHisMetLysAsnPheValThrLysValSerValG 223
598 GGAGAGTGGCCGCCACACATGAAGAATTTGTGACCAGAGGTTTCAGTTG 647
223 lYgluPheValGlyGluGlyGlyLysSerLysLysIleSerLysLys 239
648 GGGAGTTTGTAGGGGAAGAGAAGGAAGAAAGCAAGAGATCTCCAGAAG 697
240 AsnAlaAlaIleAlaValLeuGluGluLeuLysLysLeuProProLeuPr 256
698 AATGGCGCAAGGCGCGTGTGAGACAGCTAAGAGAGGCTGCCACCTTACC 747
256 oAlaValGluArgValLysProArgIleLysLysLysThrLysProIleV 273
748 TGCTGTGAGCGAGTGAAGCCCAAGATCAAGAAAGTCAAGCCCACT 797
273 aLysProGlnThrSerProGluTyrGlyGlnGlyIleAsnProIleSer 289
798 GCAAGCTACAGACTGCCCGCCGACTATAGGCCAAGGAGTGAATCTATTAGC 847
290 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrTh 306

|||||
848 AGACTGGCAGATCCACGACGCAAAAAAGAGAGAGCCGAGTACAT 897
306 rleuleuThrGluArgGlyLeuProArgArgGluPheValMetGlnv 323
898 GCTCCTTACAGAACGAGGCTCTCCACGTCGCGAGGAGTTGTGATGCAGG 947
323 allysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLysVal 339
948 TGAAGGTGGGCATCACACCGCAGAGAGAGGAGTACCAATTAAGAAGGTG 997
340 AlAlAspArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysVa 356
998 GCCAAGCGCAATGCTGCTGAAGAACATGCTGAGATCCTGGGTTCAAGT 1047
356 lProGlnArgGlnProThrLysProAlaLeuLysSerGluLulysThrp 373
1048 CCCCCAGCGCAGCCGCGCAAGCCAGCAGCTCAATCAAGAAGAGACTC 1097
373 rolleLysLysProGlyAspGlyArgLysValThrPheGluProGly 389
1098 CAGTAAAGAACCAGAGAGATGGAAGAAAGTACATTTTGAACCTAGC 1147
390 SerGlyAspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetPr 406
1148 CTTGGGGATGAAAATGGAACCACTAATAAGAGAGATGAGTTCAGAGATGCC 1197
406 oTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValProG 423
1198 CTATCTTAGCATCAGACAGCTGCCAGCCGGGATTCCTCCCATGTTGCCAG 1247
423 luValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPhe 439
1248 AGGTGGCCAGGCTGCGGGGTTAGTCAAGGACACCATACCAAGATTC 1297
440 ThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAl 456
1298 GCCAGGGCAGCTCCAAATCCTGCCAAGGCCAGCGTAACCTGCATGATAGC 1347
456 aArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuL 473
1348 CCGAGATTTGTTGACGGGGGCACCTCGCCACAGCCGAGACCATTTTAA 1397
473 ysAsnAsnIleSerSerGlyHisValProHisGlyProLeuThrArgPro 489
1398 AGAGTAACATCTCTTACAGACATGTACCCCATGGAACCTCGCACAAAGACC 1447
490 SerGluGlnLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTy 506
1448 TCTGAGCAACTGTACTACTTTCAGAGCCGAGGATTCAGGTTGAATA 1497
506 rLysAspPheProLysAsnLysAsnGluPheValSerLeuIleAsn 523
1498 CAAAGACTTCCCAAGAACAAAGAGAGAGTGTATCTCTCATCAACT 1547
523 ysserSerGlnProProLeuIleSerHisGlyIleGlyLysAspValGlu 539
1548 GCTCTTCACAGCCACTCTCGTCAGCCATGGCATCGGCAAGATGTGAG 1597
540 SerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLe 556
1598 TCCTGTCAATGATATGGCTGCACCTGAACATCTTAAAGCTGTCTGAGTT 1647
556 uaspgInGlnSerThrGluMetProArgThrGlyAsnGlyProMetSerV 573
1648 GCACCAACAGACAGACAGATGCCAAGGACAGAAACGACACAGTTTCAG 1697
573 aICysGlyArgCys 577
1698 CGTGGCGGAGGTGC 1711
seq_name: gb_ro:AF061942

seq_documentation_block:
LOCUS AF061942 2860 bp mRNA ROD 04-MAR-1999
DEFINITION Mus musculus RNA-binding protein stauflen mRNA, complete cds.
ACCESSION AF061942
VERSION AF061942.1 GI:4335944
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2860)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.
TITLE Mammalian stauflen is a double-stranded-RNA- and tubulin-binding
JOURNAL Mol. Cell Biol. 19 (3), 2220-2230 (1999)
MEDLINE 99147057
REFERENCE 2 (bases 1 to 2860)
AUTHORS Duchaine,T., Luo,M. and DesGroselliers,L.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES
source 1..2860
/organism="Mus musculus"
/db_xref="taxon:10090"
326..1789
/codon_start=1
/product="RNA-binding protein stauflen"
/protein_id="AADI7529.1"
/db_xref="GI:4335945"
translation="MYKPVDPDHSRMOSTYSYGMRGAYPRFYFPVPVPLLYQVELS
VGGOQFNKGKMRPVPKHDAPARALRTLOSEPLPERLEVNGREAEEENLNKSEISQVF
EIALKRLNPVNEFVARESGPRHMKNFVTRVSGEFGEGESKSKISKNAARAVLEQ
LRRLPPLPAVERVKPRIKKSQPTCKTAPDYGGQNPISRLAQIQAKKEPEYML
TERGLPRRREFVMQVKVGHHTAEGVTNKVAKRNAENMLEILGFVPOAQPAKRAL
KSEETPVKPGDGRKVTFFPSPGDENGTSNKDEFPMPLYSHOQLPAGILPMVPEV
AQAVGVSGHHTKDFTRAAPPAKATVTAMIARELLYGTSPTAETILKSNISSGHVP
HGPRTRPSEQLYLSRAOGFQVEYKDEPKNNKNECVSLINCSSOPPLVSHGICKDVES
CHDMAALNIKLSELDDQSTEMPRTGNGPVSACGRC"
BASE COUNT 747 a 750 c 713 g 650 t
ORIGIN
alignment_scores:
Quality: 2422.50 Length: 552
Ratio: 4.750 Gaps: 2
Percent Similarity: 92.391 Percent Identity: 85.145
alignment_block:
us-09-316-048-6 x AF061942 ..
Align seg 1/1 to: AF061942 from: 1 to: 2860
26 SerLeuLeuSerGlnProLeuMetSerIleProSerThrThrSerSerle 42
|||||
158 TCCCTCCTCTGTCGCGCCCTTTTCTCCGCGCTTCACCTGCTCTTC 207
42 uProSerGluAsnAlaGlyArgProIleGlnAsnSerAlaLeuProSera 59
|||
208 ACCTCCTCGCGCGGCCCAAGA.....CCGCCGG 236
59 laseRIleThrSerThrSerAlaAlaAlaGluSerIleThrProThrVal 75
||
237 CCCCCGACGAGCTCTGGGGAAGCAGCCAGAAAGTATAGCTTACCATTT 286
76 GluLeuAsnAlaLeuGlyMetLysLeuGlyLysLysProMetTyrLysPr 92
|||||
287 GAGCTCAATGCACCTGTGTGAACCTGGAAGAAACCAATGTATAAGCC 336
92 oValAspProTyrSerArgMetGlnSerThrTyrAsnTyfAsnMetArgg 109
|||||
337 CGTGACCCCTCACTCTCGGATGCAGTCCACCTACAGCATATGGCATGCGTG 386

481 ValProHisGlyProLeuThrArgProSerGluGlnLeuAspTyrLeuSe 497
||| ::|||::: ::|||:::|||||:::|||||:::|||||:::
1570 CCCCTTGTCTCCAGTA...CAACCTTCAAAACAACGATATTACG 1616

497 rArgValGlnGlyPheGlnValGluTyrLysAspPheProLysAsnAsnL 514
:||||:|||||:|||||:|||||:|||||:|||||:|||||:
1617 AAGCATTCAGGCTTTCAGGTATGAATTAAGCAAAACAAACAA 1666

514 ySAsnGluPheValSer 519
::|||:::|||||:
1667 ACAATTTCATTAGCCTCA 1683

seq_name: gb_pr4:AK002152

seq_documentation_block:
LOCUS AK002152 2968 bp mRNA PRI 22-FEB-2000
DEFINITION Homo sapiens CDNA FLJ11290 fis, clone PLACE1009622, weakly similar
to MATERNAL EFFECT PROTEIN STAUFIN.
ACCESSION AK002152
VERSION AK002152.1 GI:7023857
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
clone:PLACE1009622.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Magatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuko,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2968)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source
1.2968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1009622"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="cloning vector: pME18SFL3"
255.1967
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA92111.1"
/db_xref="GI:7023858"
/translation="MANPKEKTAMCLVNELARFNRVQPYKLLNERGPAHSMFVQL
SLGEQTWESESSIRKAAQAVANKALTESTLPKPVOKPKSVNNNPSTPTVELNG
LAMKRGEPALIRPLDPKPPNRYRANYNFRGMYNQRYHCPVPKIFYVLTVGNNEFEG
GKTROAARHNAMKALQALONEPIPEPONGESGKMDDDDKANKSEISLIVEIALK
RMPVSEFVTKESGPPHMKSFVTRVSGEFSAEGEKSKLSKRAATTVLQELKLP
PLPVEKPKLFEKKRPKTIYKAGPEYGGMPISRLAQIQAKKEKEDVLLSERGM
PRRREFVMQYVGNVATGTGPNKTIARKNAEAMLLQLGKASTNLQDLEKTEGENK
GMSGPKPGFPEPTNNTPKGILHLSPDVYQEMEAARHKVISGTTIGLYLSPKDMNPSSS
FFSISPTSNSSATIAELLMNGTSTFAEIGLKSSPTPCSPVQPSKOLEYLARIQG

FOAALSALKQFSEQGLDPIDGAMNIEKSLKQAKHLREKADNNQAPPGSIAQDCCKS
NSAV"
BASE COUNT 970 a 605 c 643 g 750 t
ORIGIN

alignment_scores:
Quality: 1067.50 Length: 523
Ratio: 2.885 Gaps: 19
Percent Similarity: 70.746 Percent Identity: 49.331

alignment_block:
US-09-316-048-6 x AK002152 ..

Align seg 1/1 to: AK002152 from: 1 to: 2968

1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG 17
||| ::|||:::|||||:::|||||:::|||||:::|||||:::
369 ATGTTCTCAGTGCAGCTGAGTCTTGTTGAGCAGACATGCGAATCCGAAG 418

17 ySerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeuMets 34
||||:|||| ::|||:::|||||:::|||||:::|||||:::|||||:::
419 CAGCAGTATAAGAAAGGCTCAGCAGCGCTGTTGCCAATAAGCTTTG... 464

34 erIleProSerThrThrSerSerLeuProSerGluAsnAlaGlyArgPro 50
||| |||:::|||||
465ACTGAATCTACGCTTCCC.....AAACCA 488

51 lIleGlnAsnSerAlaLeuProSerAlaSerIleThrSerThrSerAlaAl 67
::|||::: |||:::|||||:::|||||:::|||||:::|||||:::
489 GTTCAGAGACCA.....CCCAAAAGTAATGTTAACAATAACCCAGGC.. 530

67 aAlaGluSerIleThrProThrValGluLeuAsnAlaLeuGlyMetLysL 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
531AGTATACTCCCAACTGTGGAAGTGAATGGGCTTGCTATGAAAA 573

84 euGlyLysLysProMetTyrLysProValAspProTyrSerArgMetGln 100
||||: ::|||:::|||||:::|||||:::|||||:::|||||:::
574 GGGAGAGCGCTGCCATCTACAGGCCATTAGATCCAAAGCCATTCCCAAT 623

101 SerThrTyrAsnTyrAsnMetArgGlyAlaTyrProArgTyrPh 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
624 TATAGAGCTAATTACAACCTTTGCG...GGCATGTACAATCAGAGGTATCA 670

117 eTyrProPheProValProProLeuLeuTyrGlnValGluLeuSerValG 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
671 T.....TGCCCAAGTGCCCTAAGATCTTTAT...GTTACGCTCAGTGTAG 711

134 lYgLYGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaLysHis 150
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
712 GAATATATGATTTTGTGGGAAGGAAGAAGACTGCACAAGCTGCTAGACAC 761

151 AsPAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGl 167
::|||::: |||:::|||||:::|||||:::|||||:::|||||:::
762 AATGCTGCAATGAAGCCCTCCACAGCACTGCAGAAATGAACCTATTCCAGA 811

167 uArgLeuGluValAsnGlyArgGlu.....SerGluGluGluA 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
812 AAGATCTCTCTCAGAAATGTTGAATCAGGAAGATATGATGATGACAAAG 861

180 snLeuAsnLysSerGluLysSerGlnValPheGluIleAlaLeuLysArg 196
::|||::: |||:::|||||:::|||||:::|||||:::|||||:::
862 ATGCAATATAGCTGAGATCAGCTTAGTGTGTAATTCGCTGAAGCGA 911

197 AsnLeuProValAsnPheGluValAlaArgGluSerGlyProProHisMe 213
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
912 AATATGCCCTGTCAGTTTGGAGGTTATTAAAGAAAGTGACCAACACATAT 961

213 tLysAsnPheValThrLysValSerValGlyLupheValGlyGlyG 230
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
962 GAAAAGCTTTGTTACTCGAGTGTCAAGAGAGAGTCTCTCGAGAAAGAG 1011

230 lUGlyLysSerLysLysIleSerLysLysAsnAlaAlaAlaValLeu 246

1012 AAGGAATAGCAAAAACCTCTCCAGAGCGCGCTGCGACCACCGCTCTTA 1061
247 GluGluLeuLysLysLeuProProLeuProAlaValGluArgValLysPr 263
1062 CAGGAGCTTAAAAAACTTCCACCTCTCTGTTGGTGGAAAAAGCCAAAAC 1111
263 oArgIleLysLysThrLysProIleValLysProGlnThrSerProG 280
1112 ATTTTAAAAAACGCCCTAAACAATAGTAAAG....GCCGACCAG 1155
280 LuTyrgLysGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGln 296
1156 AATATGCCAAGGATGAACCTATTAGCCGCGCTGGCCAAATTCACAG 1205
297 AlaLysLysGluLysGluProGluTyThrLeuLeuThrGluArgGlyLe 313
1206 GCCAAAAGGAAAAGAGCGCGGATATGTTTGGCTTCAGAAAGAGAAT 1255
313 uProArgArgArgGluPheValMetGlnValLysValGlyAsnHisThra 330
1256 GCCTCGACGTCGAGAATTGTGATGACGTGAAGGTAGGCAATGAAGTTG 1305
330 laGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaGlu 346
1306 CTACAGGAACAGACCTAATAAAAAAGATAGCCAAAAAATGCTGCAGAA 1355
347 AsnMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProThrLy 363
1356 GCAATGCTGTACAACTTGTTATAAAGCATCCACTAAT..... 1394
363 sProAlaLeuLysSerGlu...GluLysThrProIleLysLysProGlyA 379
1395CTTCAGATCAACTTGTTATAAAGCATCCACTAAT..... 1437
379 sPGLyArgLysValThrPheGluProGlySerGlyAspGluasnGly 395
1438 GTGGTCCAAAGCCGTGGGTTTCCTGAACCA..... 1466
396 ThrSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisGlnGl 412
1467 ACAATAAT..... 1475
412 nLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaVal. 428
1476 .ACTCCAAAAGAAATTCTTCATTGTCCTCGATGTTTATCAAGAGATGG 1524
429GlyValSerGlnGlyHis...His 435
1525 AAGCCAGCCGCCACAAAGTAATCTGCGACTACTCTAGCTATTGTCA 1574
436 ThrLysAspPheThrArgAlaAlaProAsn.....ProAlaLys 448
1575 CCCAAAGATATGAACCAACCTTCAAGCTCTTCTTCAGTATATCTCCAC 1624
448 salatThrValThrAlaMetIleAlaArgGluLeuLeuTyrgLysThrs 465
1625 ATCGAATAGTTCAGCTACAAATTCGCCAGGGAACCTCTTATGAATGACAT 1674
465 erProThrAlaGluThrIle..LeuLysAsnAsnIleSerSerGlyHis 480
1675 CTTCCTACAGCTGAACCATAGGTTTAAAGA....AGTTCCTCTACT 1718
481 ValProHisGlyProLeuThrArgProSerGlnGlnLeuAspTyrLeuSe 497
1719 CCCCCTGTGTTCCAGTA...CAACCTTCAAAAACAACCTGAATATTAGC 1765
497 rArgValGlnGlyPheGln 503
1766 AAGGATTCGAAGCTTTCAG 1784
seq_name: gb_htg14:AC063967

seq_documentation_block:
LOCUS AC063967 225016 bp DNA 06-SEP-2000
DEFINITION Mus musculus clone RP23-41G23, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC063967
VERSION AC063967.3 GI:9972323
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 225016)
REFERENCE MCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Shekher,M., Spiegel,L.A., Toth,K. and Vill,M.D.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 225016)
REFERENCE MCombie,W.R.
Direct Submission
Submitted (22-APR-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Sep 6, 2000 this sequence version replaced gi:8439872.
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Center project name: RP23-41G23
Center clone name: RP23-41G23

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 121343: contig of 121343 bp in length
* 121344 121985: gap of unknown length
* 121986 139930: contig of 17945 bp in length
* 139931 140572: gap of unknown length
* 140573 155472: contig of 14900 bp in length
* 155473 156114: gap of unknown length
* 156115 167728: contig of 11614 bp in length
* 167729 168370: gap of unknown length
* 168371 177185: contig of 8815 bp in length
* 177186 177827: gap of unknown length
* 177828 185945: contig of 8118 bp in length
* 185946 186587: gap of unknown length
* 186588 194327: contig of 7740 bp in length
* 194328 194969: gap of unknown length
* 194970 202341: contig of 7372 bp in length
* 202342 202983: gap of unknown length
* 202984 210036: contig of 7053 bp in length
* 210037 210677: gap of unknown length
* 210678 216586: contig of 5909 bp in length
* 216587 217227: gap of unknown length
* 217228 220778: contig of 3551 bp in length
* 220779 221419: gap of unknown length
* 221420 223928: contig of 2509 bp in length
* 223929 224569: gap of unknown length
* 224570 225016: contig of 447 bp in length.

FEATURES
source
1. 225016
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-41G23"

BASE COUNT 59632 a 50195 c 50037 g 57445 t 7707 others
ORIGIN

alignment_scores:

Quality: 921.00 Length: 1792
Ratio: 2.544 Gaps: 11
Percent Similarity: 20.201 Percent Identity: 18.694

alignment_block:

US-09-316-048-6 x AC063967/rev ..

Align seg 1/1 to reverse of: AC063967 from: 1 to: 225016

```
203 GluValAlaArgGluSerGlyProProHisMetLysAsnPheValThrLy 219
      ::::::::::::::::::::::::::::::::::::::::::::::
89063 CAGGTGGCCCGGAGAGAGTGGCCACCACACATGAGAACTTTGTGACCAG 89014

219 sValSerValGlyGluPheValGlyGlyGlyLysSerLysLysI 236
      ::::::::::::::::::::::::::::::::::::::::::::::
89013 GGTTCAGTTGGGGAATTTGAGGGAGAGAGAGGAAAGCAAGAAGA 88964

236 lSerLysLysAsnAlaAlaIleAlaValLeuGluLeuLysLysLeu 252
      ::::::::::::::::::::::::::::::::::::::::::::::
88963 TCTCCAAGAAGATGCGCCAGGGCTGTTCTGAGCAGCTTAGGAGGCTG 88914

253 ProProLeuProAlaValGluArgValLysProArgIleLysLysLysTh 269
      ::::::::::::::::::::::::::::::::::::::::::::::
88913 CCACCCCTCCCTGCTGTGTGAGCGAGTGAAGCCCAAGATCAAGAAGAAAG 88864

269 rLysPro..... 271
      :::::
88863 TCAGCCCACTGCAAGGTGAGAGCCCTTCCCAGAAACCACAGGCAAGG 88814

271 ..... 271
88813 AGAGAGCCGGCTGCTGTGTGGGTGGGTAGAAAGCTGCTCAGCACC 88764

271 ..... 271
88763 ACAGGGGACAGAATGATGTTTACAGACAGACATAGAGCTATAGTTG 88714

271 ..... 271
88713 AGATAAATGAGCCACTGATGGGTGAGAGAGAAACGTCCAGGTGTCA 88664

271 ..... 271
88663 GCTCCCTGGGGTGGGCACACAGCCCTGCGTCTTTGTGTGAAGGG 88614

271 ..... 271
88613 GTAGGTTTCTTGAGCCAAGTCTCAGGATCAAGGTGAAGCAGCA 88564

271 ..... 271
88563 CTGAGGGCCACAGAGACTATTCCAAGACTTGATGCTTGCAATGAAC 88514

271 ..... 271
88513 TTAGTGCTATAGTTAATAATAATAACAGCTGGGCATGGTGGCGCAC 88464

271 ..... 271
88463 GCCTTTAATCCAGCACTTGGGGGGCAGAGGCAGACAAATTTCTGAGTTC 88414

271 ..... 271
88413 GAGCCAGCCTGCTCTACAGATGAGTTCCAGAGACAGCCAGGGCTACACA 88364

271 ..... 271
88363 GAGAAACCTGTCTCGGAAAAAATAAAAAATACAGCATTCACCATGAGTA 88314
```

```
271 ..... 271
88313 CTGAGACTGGAGAGGTGGCTCAGCAGTTAAAGACTGGCTGCTAGC 88264

271 ..... 271
88263 AGAGAGCTGGGTTCAACAATCTGTACACACCAGTTCAGGACTCCTCT 88214

271 ..... 271
88213 GAACCTCCTGAGCACCAAGTGACATGTGAACCTGATGCAGACAAACAC 88164

271 ..... 271
88163 TCACATACAGAAGATAGGAATCTTATTACAGCATTGGCCCTTTCCGC 88114

271 ..... 271
88113 ACTAGAAAGGACAGAGACTTCAGCCAATCTCTCTTTTCATCTGTGTG 88064

271 ..... 271
88063 TCCTGAATAATTACTTCATCATCCCTGACAGGTGCCATGGGGAAGAAG 88014

272 ..... 275
      IleValLysPro.....
88013 AGCCCTAGTGTCTCTCAGAAGCTTGACGTCTCTCTCCCGGGCCTCACC 87964

275 ..... 275
87963 TGGAGGTTTAAATTCAATACAGTGTTTGTTTGTCTACTGGGATTG 87914

275 ..... 275
87913 GATGAGACTCAGGTGTGCTGACCAAGAGCCCTTCCATAGGATG 87864

275 ..... 275
87863 CATCTTACACTGAGGGCTTACAAAAGCTTTAAGTGTGCTGGGTTT 87814

275 ..... 275
87813 AATGATGAGAAAGCTAGCATGGTCCCATGTAGATTGATGACGACA 87764

275 ..... 275
87763 GGTCAAAGCCCTGGTTCATCCCTCTTTTGTCAATCTTTT TTTT 87714

275 ..... 275
87713 TTTTCCGGAACCTCACTCTGTAGACCAGGCTGGCCTCGAATTCAGAAAT 87664

275 ..... 275
87663 CCGCCTGCTCCACCTCCCAAGTCTGGGATTAAGGGCTGCACACCAC 87614

275 ..... 275
87613 TGCCCGGCTCCTTTTGTCAATTCGCCCAAGTGTGTCTTACAGACTTTT 87564

275 ..... 275
87563 ACATGCGAGTTCACCTGTGCTAATGTCCCAAGTTACATGAGAGTCAGCTTC 87514

275 ..... 275
87513 CGCATGCCGAGAGTGACAGAGGTCAAGGCAAAACACCAACCAAGAAGCA 87464

275 ..... 275
87463 TCTCAGCATGATCCTGAACATGTTTAGTCTTAGAGAGGGTTTGGTCT 87414
```

```
275 ..... 275
87413 GAGCGTTACACACACTAGTAGTTCTGTCAGTCCCGTGTGAC 87364
275 ..... 275
87363 GCTGACTGCTGGTTCCTCTCTGTGAGTAACAGCATATTGTCCTGCAC 87314
276 .....GlnThrSerProGluTyrGlyGlnGlyLeasn 286
      |||||:::||:|||||:::|||
87313 TGCTTGTTTTCAGCTACAGACAGCCCCGGATTATGGCCAGGATGAAT 87264
287 ProIleSerArgLeuAlaGlnIleGlnAlaLysLysGluPr 303
      |||||:::|||:::|||:::|||:::|||:::|||
87263 CCTATTAGTAGACTTGCACAGATCCAGCAGCAAAAAAGAGAGAGAGACC 87214
303 oGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgGluPheV 320
      |||||:::|||:::|||:::|||:::|||:::|||
87213 AGAGTACATGCTCCTTACAGAACGAGGCTTCCACGTCGACAGGAGTTTG 87164
320 alMet..... 321
      |||||
87163 TGATGCAGGTGAGTTGCTGGGGTTGTAGCAGCCCTTCCCTGACCATGTC 87114
321 ..... 321
87113 CTAATCCGCGAGTGTTCAGTTAGCTGACTTGGTAGTGCCCTGCTTG 87064
321 ..... 321
87063 CTCTCTGAGGCCAAGCAAGCGGAGGTGATCTCTATGAGTTCAAGAC 87014
321 ..... 321
87013 TCCACATTAGTAGATCCCATCACAGAAACAGATCCACTCTGTAAAGAT 86964
321 ..... 321
86963 GAGTTGACCAGAGTTTGCCCTGAAGTGGCTGGGGGACAGCCTTGTGAAGA 86914
321 ..... 321
86913 TGTGCTGAATTGTGCCAGCAGCCCTCAGAGAGAGTGTATGAGGG 86864
321 ..... 321
86863 AGAGCCCCACTGCACCTAGATATATACCTGTCTTAATAATTGATGAT 86814
321 ..... 321
86813 GTGTTGTGGACACACTTTCCATACAATAGTTTAAAGTTAAAGAAG 86764
321 ..... 321
86763 TCTTTAAGACTGTGAGATGGCTCAGTGGTTAAGAGCACTGACTGCTCT 86714
321 ..... 321
86713 TCCGAAAGTCCGTGAGTTCAAAATCCAGCAGCAGCATGTGGCTGCAACC 86664
321 ..... 321
86663 ACCCATAAATGAGATCTGATGCCCTCTTCTGGTGCATCTGAAGACAGCTAC 86614
321 ..... 321
86613 AGTGTACTTAATAAAGTCTTTAAAAAAAATAAACTCTTTAAAAATAGA 86564
321 ..... 321
86563 GTTTCAATATCTCTAAGGTGTGTGGCCATCAGACAGTAGTAGTTCA 86514
321 ..... 321
```

```
86513 GAATACGAACAATAAGCTTTTACCATCAAGTTCAGCAGTCATCCCTGG 86464
322 .....GlnValLysV 325
      |||||
86463 GCCTTTGGGTTAAGTCACAGGCTCTTGTGTGGTGTTTTACACAGTAAGG 86414
325 aGlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLysValAlaLys 341
      |||||:::|||:::|||:::|||:::|||:::|||
86413 TTGGGCATCACACTGCAGAAAGAGAGTGGGTACCAATTAAGAGGTGGCCAAG 86364
342 ArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysValProG 358
      |||||:::|||:::|||:::|||:::|||:::|||
86363 CGTAATGCTGCTGAGAACATGCTGGAGATCCTGGGGTTCAAAAGTTCCCA 86314
358 nArgGlnProThrLysProAlaLeuLysSerGluLys..... 371
      |||||:::|||:::|||:::|||:::|||:::|||
86313 GCGGCAGCCTGCCAAGCCAGCACTCAAATCAGAAAGAGTAGTACTGCTG 86264
371 ..... 371
86263 GGACTCTGTGCCAGGTGGCCCTAACTCCCTGCATGGGCAAGCCTGGGCAA 86214
371 ..... 371
86213 GCCTTCATCTCCCTGCAGTAAGCGGCTGCCCGGGGCTCTCCATT 86164
371 ..... 371
86163 GAGATTAGATTGAGAGGAGGAGTTTCTGTCCAGCTCGTCACATCTTG 86114
371 ..... 371
86113 GAGCCATCATTCGGTTAAGTCTCTGTTAGTGTCACTATGGGCTAGAT 86064
371 ..... 371
86063 GTGCATCTCCGAGAGAGTCTCTGCTCTACTTCTGATGGTAGTTGGGG 86014
371 ..... 371
86013 TTGAGAGGAACATAAATGACTGCCACCTGTCTGCATGACCCAGTTTGAA 85964
371 ..... 371
85963 TATGCCCTCTGTTTCTCTGGGACAAGCTAGCATGGTCAAATAAGCCAAAT 85914
371 ..... 371
85913 CTATAGATTGAGTTGAATATCACTATACGCCACCATTTGCTGGCTACAGC 85864
371 ..... 371
85863 AGTTGCATATTGGGATTTGAGCTTCTGAGCTCTGCCAGTTTCATTACA 85814
371 ..... 371
85813 AGAATAGCTCCCTCCCACTGCTGGAGCTTAGTACTTGTAAGTGGCTT 85764
371 ..... 371
85763 GATTGGCTTCCAGATTTCACATTCCAAACAGTAGAGTCTATTAGTT 85714
371 ..... 371
85713 CAAGACTCCACATAGTAGATCCCGTACAGAAACAGTATCCGCTCTGTA 85664
371 ..... 371
85663 AGAGTAGTTGACCAAGAGTTTGCTGAAGTGGCTGGGGACAGCTTGA 85614
371 ..... 371
```


85613 TGAAGTGTGCTGGAATCTGTCTCTTCTGTGTGATAGCATTAGGTACCCTC 85564
371 371
85563 TGTGTCTTCTTGTGGATGTACTGTTTTGGATGTCTTAGTATGTAAGTG 85514
371 371
85513 TTCTGTCTTAGCACACAGTGCAGACCCTACTTTCCACAAGCTGAACACT 85464
371 371
85463 TTTAAGCCAGCATTTTCTCTGGCAGTGGTTTCTGTCTTGGCACAGT 85414
371 371
85413 GCCAGCTATGCATTCAGATCTGCTGTGCAAAATGCTCCATGGGAGG 85364
371 371
85363 CGCAGCGTTTGTCTATCCACAGACCCAGTGTCTGTATGCTCACT 85314
372 ThrProIleLysLysProGlyAspGlyArgL 382
85313 GCCATTTTCTTCTTAAAGACTCCAGTAAGAAACCAAGAGACGGAAGA 85264
382 ysValThrPhePheGluProGlySerGlyAspGluAsnGlyThr..... 396
85263 AAGTAACGTTTTTTGAACCTAGCCCTGGGATGAATAAGAACTAGTAAG 85214
396 396
85213 TGTGCCCTCTTGTCAATTGAACCAATGTGCTGTGTCTGTGTGGCTCT 85164
396 396
85163 CGTGAGTTTTTGGAAGCACAGATATCTCTGCCCTTCTGTATGCTGGGAT 85114
396 396
85113 GGTAAAGGTGCTCTGTAGCTTCTGGGTGGTGAATGCTTGTGGTGAC 85064
397 Ser..AsnLysGluAspGluPheArgMetProTyrLeu 408
85063 CTGTGTCCCTGCAGGTAAACAAGACGAGAGTTCAGATGCCCTTATCTT 85014
409 SerHisGlnGlnLeuProAlaGlyIleLeuPrometValProGluValAl 425
85013 AGCCATCAGCAGCTGCCAGCTGGAATTCCTCCCATGTGCGCGAAGTTGC 84964
425 agLAlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArg 442
84963 CCAGGCTGTGGGGTTAGTCAAGACACACACCAAGATTTCACACAGGG 84914
442 laAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 458
84913 CAGCTCCAATCCTGCCAAGCAACGTTAACTGCCATGATAGCCCGAGAG 84864
459 LeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAs 475
84863 TTGTGTACGGGGGCACCTGCCACACAGCCGAGACCATTAAAGAGTAA 84814
475 nIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGluG 492
84813 CATCTCTTCAGGCCACAGTACCCCATGGAACCTGCACCTAGACCCCTGAGC 84764
492 InLeuAspTyrLeuSerArgValGlnGlyPhe.Gln..... 503
84763 AACTGTACTACTTTCCAGAGCCAGGGATTCAGAGTTAACCGTGTGTCA 84714
503 503
84713 TCTCTGGGCTTACGTTCTTCCTCACAGAAATTGCTTGTAGTGTGTGC 84664

503 503
84663 AAGACCCGTGGTCTGTCCCTCAGCACTACAGAAACCAGAAAGCCAAGAT 84614
503 503
84613 TATATGCTTATTTAAGATTGATGTACAGTGAAGTTCGAAGAAGGA 84564
503 503
84563 AGATGTCTGGGATGAGAATGGCTTATATAAGGTAGCTATTCTAAGCCC 84514
503 503
84513 AACAACTTTAGTATGATTACTGTGGAAGAAACCAACTCCAGCAAGTTG 84464
503 503
84463 TCCTCAGTGTGCACACCCTGCCCGAGAGTTTACGTGTGCCAGGCTG 84414
503 503
84413 AAGGAGAGAAACACTTTTACTTAAAGAACACCCACTACATGACATGG 84364
503 503
84363 GAGAAGAGCTTGAGGGTCTGTGGGCTGGGCCCATATGCAAGGTTGCT 84314
504 ValGluTyrLysAspPheProLysAsnLysAsnGlu 516
84313 GTTTTCTCTAGGTTGAATACAAAGATTTTCCCAAGAACCAAGAAGCGA 84264
516 upheValSerLeuIleAsnCysSerSerGlnProProLeuIleSerHisG 533
84263 GTGTGATCTCTCATCACTGCTCTCCACAGCCGCTCTGCTCAGTCATG 84214
533 LylleGlyLysAspValGluSerCysHisAspMet..... 544
84213 GCATCGGCACAGATGTGAGTCTCTCATGATATGATGTGGGCCCAGG 84164
544 544
84163 AGGTGTGAGAAGAGCTGGCACATTCCTTGTTCAGTCAGTGTAGTG 84114
544 544
84113 GCCTCAGAAAGAGGACATGTTCTCTGTAGCCTTGTGACATGTCTGA 84064
544 544
84063 TTGATTCATTCAATTCTCTCTCTGCTTCTTCTCTCTCTCACTTCT 84014
544 544
84013 GGCAGCTGTAGAAAGCCCCCTTGGGAGTTCTGTTCAAAGAAAGAAATCA 83964
544 544
83963 GCACACTTGAGCTTTGGTGTTCACCTGATGGAGCATTAATGCACCCCTTTA 83914
544 544
83913 AGCGGACACCTCTGATTTGAAGGAAACATGGAGTTTGATGTGAGCA 83864
544 544
83863 GGATTCAAACATTTTCATTGGGAAGTTTAAAGAGCTGAGCTGCTGTTC 83814
545 Ala..AlaLeuAsnIle 549
83813 CCTGTGTTGATTATTCCTCTCTCACTTCTTAGGCTGCACCTGAACAT 83764

549 eLeuLysLeuLeuSerGIuLeuAspGIuInserThrGIuMetProArgT 566
|||||
83763 TTTAAAGCTGCTGTCGTGAGTTGGACCAACAGACAGACAGATGCCAAGAA 83714
566 hriGIyASnglyPrometSer 572
|||||
83713 CAGGAAATGACACAGTTTCA 83694
seq_name: gb_pr6:HSBB140E4
seq_documentation_block:
LOCUS HSBB140E4 163253 bp DNA PRI 07-FEB-2000
DEFINITION Human DNA sequence from clone RP13-140E4 on chromosome X, complete
sequence.
ACCESSION AL121877
VERSION AL121877.13 GI:6933862
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163253)
AUTHORS Heath,P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Feb 7, 2000 this sequence version replaced gi:6911935.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP13-140E4 is from the library RPCI-13.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: PBACE3.6
This sequence is the entire insert of clone RP13-140E4.
FEATURES
Source
1.163253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-140E4"
/clone_lib="RPCI-13.1"
25.188
repeat_region
/note="L1M4 repeat: matches 5114. .5296 of consensus"
184.411
repeat_region
/note="L1P3 repeat: matches 2841. .3072 of consensus"
412.694
repeat_region
/note="AluSq repeat: matches 1. .287 of consensus"
695.3450
repeat_region
/note="L1P3 repeat: matches 232. .2841 of consensus"
3451.6427
repeat_region
/note="L1PA2 repeat: matches 3148. .6144 of consensus"
6432.7143
repeat_region
/note="L1MD1 repeat: matches 5283. .6016 of consensus"

repeat_region
7164. .7316
/note="AluJb repeat: matches 137. .290 of consensus"
7339. .7646
repeat_region
/note="L1MD2 repeat: matches 6016. .6330 of consensus"
7650. .7754
repeat_region
/note="L1MC5 repeat: matches 7608. .7714 of consensus"
7767. .8072
repeat_region
/note="AluY repeat: matches 1. .305 of consensus"
8095. .8136
repeat_region
/note="21 copies 2 mer ca 90% conserved"
8152. .8245
repeat_region
/note="L1MC5 repeat: matches 7478. .7572 of consensus"
8273. .8494
repeat_region
/note="AluJo repeat: matches 84. .305 of consensus"
8503. .8633
repeat_region
/note="L1MC5 repeat: matches 7305. .7446 of consensus"
8712. .8879
repeat_region
/note="84 copies 2 mer tt 72% conserved"
8899. .9190
repeat_region
/note="AluJo repeat: matches 8. .300 of consensus"
/complement(9869. .10121)
misc_feature
/note="match: STS: Em:L41960; match: STS: Em:L41960"
12531. .13017
repeat_region
/note="L2 repeat: matches 2083. .2505 of consensus"
13018. .13320
repeat_region
/note="L1MA1 repeat: matches 6004. .6303 of consensus"
13321. .14004
repeat_region
/note="L2 repeat: matches 1197. .2083 of consensus"
/complement(16833. .17221)
misc_feature
/note="match: GSS: Em:AQ558019"
16840. .17277
repeat_region
/note="match: GSS: Em:B65742"
/complement(16845. .17241)
misc_feature
/note="match: GSS: Em:AQ149727"
/complement(16848. .17220)
repeat_region
/note="match: GSS: Em:AQ240745 Em:AQ474471"
16886. .17629
repeat_region
/note="MER4D repeat: matches 234. .973 of consensus"
18183. .18407
repeat_region
/note="L1PB3 repeat: matches 5928. .6148 of consensus"
18455. .19002
repeat_region
/note="MER41A repeat: matches 1. .554 of consensus"
19289. .19465
repeat_region
/note="MIR repeat: matches 15. .198 of consensus"
19583. .20000
repeat_region
/note="match: GSS: Em:AQ017861"
19875. .20274
repeat_region
/note="MER57B repeat: matches 1. .403 of consensus"
20275. .21084
repeat_region
/note="MER57-internal repeat: matches 1. .806 of consensus"
21078. .22422
repeat_region
/note="MER57-internal repeat: matches 440. .2799 of
consensus"
22425. .22764
repeat_region
/note="MER57-internal repeat: matches 3892. .4230 of
consensus"
22765. .23041
repeat_region
/note="AluXx repeat: matches 8. .286 of consensus"
23042. .24919
repeat_region
/note="MER57-internal repeat: matches 4230. .6114 of
consensus"
24939. .25118
repeat_region
/note="L1PB1 repeat: matches 5980. .6153 of consensus"
25119. .25422
repeat_region
/note="AluY repeat: matches 1. .304 of consensus"
25423. .25607
repeat_region
/note="L1PB1 repeat: matches 5777. .5980 of consensus"
25635. .25672
repeat_region
/note="19 copies 2 mer ta 78% conserved"
25681. .26070
repeat_region
/note="L1PB1 repeat: matches 5397. .5790 of consensus"
26082. .27136
repeat_region
/note="MER57-internal repeat: matches 6489. .7537 of

consensus"
 repeat_region 27137..27249 /note="MER57B repeat: matches 1..114 of consensus"
 repeat_region 27250..28097 /note="LIMC3 repeat: matches 6763..7631 of consensus"
 repeat_region 28123..28390 /note="MLJ1C repeat: matches 28..313 of consensus"
 repeat_region 28404..28952 /note="LIMC3 repeat: matches 5376..5993 of consensus"
 repeat_region 28946..29622 /note="LIM4 repeat: matches 4880..5566 of consensus"
 repeat_region 30851..31163 /note="ALUSC repeat: matches 1..301 of consensus"
 repeat_region 32518..34668 /note="LIPAL4 repeat: matches 3983..6149 of consensus"
 repeat_region 34672..35189 /note="MLT2B repeat: matches 1..440 of consensus"
 repeat_region 35190..36499 /note="LIPAL4 repeat: matches 2718..3988 of consensus"
 repeat_region 36500..36876 /note="THB1C repeat: matches 1..369 of consensus"
 repeat_region 36877..38213 /note="LIPAL4 repeat: matches 1264..2718 of consensus"
 repeat_region 38221..39169 /note="LIPAL4 repeat: matches 288..1240 of consensus"
 repeat_region 39455..39517 /note="L2 repeat: matches 2349..2404 of consensus"
 repeat_region 40251..40337 /note="MIR repeat: matches 59..146 of consensus"
 repeat_region 41354..41648 /note="ALUSX repeat: matches 1..295 of consensus"
 repeat_region 41650..41743 /note="47 copies 2 mer aa 62% conserved"
 repeat_region 41744..42318 /note="L1 repeat: matches 4553..5128 of consensus"
 repeat_region 42337..42391 /note="ALU repeat: matches 251..302 of consensus"
 repeat_region 42360..42395 /note="ALU repeat: matches 79..114 of consensus"
 repeat_region 42421..43734 /note="L1 repeat: matches 4019..5334 of consensus"
 repeat_region 43735..43834 /note="FLAMC repeat: matches 32..131 of consensus"
 repeat_region 43835..44621 /note="L1 repeat: matches 3232..4019 of consensus"
 repeat_region 44622..44802 /note="L1M1 repeat: matches 6121..6304 of consensus"
 repeat_region 44803..45815 /note="L1 repeat: matches 2126..3232 of consensus"
 repeat_region 45514..46762 /note="L1M1 repeat: matches 1155..2794 of consensus"
 repeat_region 46760..47327 /note="L1M1 repeat: matches 59..642 of consensus"
 repeat_region 47325..48451 /note="L1M1 repeat: matches -1334..-154 of consensus"
 repeat_region 48533..48727 /note="L1 repeat: matches 4030..4226 of consensus"
 repeat_region 48717..49410 /note="L1M9 repeat: matches 5092..5805 of consensus"
 repeat_region 49411..49506 /note="48 copies 2 mer at 69% conserved"
 repeat_region 49528..50023 /note="L1M9 repeat: matches 5787..6308 of consensus"
 repeat_region 50050..50333 /note="ALU repeat: matches 1..300 of consensus"

alignment_scores:
 Quality: 798.00 Length: 536
 Ratio: 2.242 Gaps: 21
 Percent Similarity: 66.418 Percent Identity: 42.537

alignment_block:
 US-09-316-048-6 x HSBBI40E4 ..

Align seg 1/1 to: HSBBI40E4 from: 1 to: 163253
 1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG 17
 158900 ATGTTCTCAGTGCACACTGAGTCTTGTTGAGCAGACATGGAATCCGAAG 158949
 17 ySerGlnLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeuMets 34
 158950 CAGCAGTATTAAGAAGGCTCAACAGGCTGTCGTTAATAAGCTTTG... 158995
 34 erileProSerThrThrSerSerLeuProSerGluAsnAlaGlyArgPro 50
 158996ACTGAATCTACGCTTCCC.....AAACCT 159019
 51 ILeGlnAsnSerAlaLeuProSerAlaSerileThrSerThrSerAlaAl 67
 159020 ATTTAGAGGCCA.....CCCAAAAGTAATGTTAACAATAACCCAGGCTG 159063
 67 aAlaGluSerileThrProThrValGluLeuAsnAlaLeuGlyMetLysL 84
 159064 T.....ATAACACCAACTGTGGAACCTGAATGGCTTGCTATGAATA 159104
 84 euGlyLysLysProMetTyrlsProValAspProTySerArgMetG1 100
 159105 GGGCAGAGCCTGCCATCCACAGGCCATTAGATCCAAAGCCATTCCCAAA 159154
 100 nSerThrTyAsnTyAsnMetArgGlyGlyAlaTyProProArgTyRP 117
 159155 TAATAGAGCTAATTACAACTTTCAA...GTCATGTATTAATCAGAGGTATC 159201
 117 heTyProheProValProProLeuLeuTyGlnValGluLeuSerVal 133
 159202 ATTGCCCAATA.....CCTAAGATCTTTTATGTTCAAGCTCAT.GTA 159241
 134 GlyGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysH 150
 159242 GGAAATATGAATTTTGGGGGAGAGGAAAGCC.....AGACA 159278
 150 sAspAlaAlaAlaLysAlaLeuArgileLeuGlnAsnGluProLeuPro 167
 159279 CAATGCTGCAATGAAGAAGACATCCAGCACTGCAGATTGAACCTATTCCAG 159328
 167 luArgLeuGluValAsnGlyArgGlu.....SerGluGluGlu 179
 159329 AAAGATCGCCTTAGAATGTGTAATCAGAAAGATATAGATGACAAA 159378
 180 AsnLeuAsnLysSerGluLeuSerGlnValPheGluLeuLeuLysAr 196
 159379 GATGCAAAATAGCCTGAGATCAGCTTAGTGTGAAATGCTCTGAAGCTG 159428
 196 gAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProHism 213
 159429 AAATATACCTGTTAGTTCTGAGGTTATTAAAGAAAGTGACACACACATA 159478
 213 etLysAsnPheValThrLysValSerValGlyGluPheValGlyGly 229
 159479 TGAAGAGCTTGTCTCTCGAGTGTGAGGAAAGTTCTCTACAGAAAGA 159528
 230 GluGlyLysSerLysLysLysSerLysLysAsnAlaAlaAlaValle 246
 159529 GAAGGAAATAGCAAAAACCTCCACAGCGTGTGCAACACCGCTCTT 159578
 246 uGluGluLeuLysLysLeuProProLeuProAlaValGluArgValLysP 263
 159579 ACAGAGCTTAATAAAACCTCCACCTCTCTGTTGGAAGCAAAAT 159628
 263 roArgileLysLysLysThrLysProileValLysProGlnThrserPro 279
 159629 TATTTTAAATAAAGCCCTAAACATAGTCAAG.....GCCGAGACA 159672
 280 GluTyGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGlnG 296

159673 GAATA.GGCCAAGGATGAACCCCTATTAGCTGCCCTGGCGCAATTCACA 159721
296 nAlAlysLysGluLysGluProGluTyrThrLeuLeuThrGluArgGlyL 313
|||||
159722 GCCCAAAAGGAAAGAGCCAGATTATGTTGCTTTCAGAAAGAGGAA 159771
313 euProArgArgArgGluPheValMetGlnValLysValGlyAsnHisThr 329
:::|||||
159772 TGCCCTCGATGTCGAGAAATTGATGATATGAGCTAGCAATGAGCT 159821
330 AlAGLUGLyThrGlyThrAsnLysLysVal...Ala.LysArgAsnAlaA 345
|||
159822 GCTACAGGAGGAGGAGCTAATAAAAGGCCAAAAAAGAGTGTG 159871
345 laGluAsnMetLeuGluLeuLeuGlyPheLysValProGlnArgGlnPro 361
||
159872 CAGTAGCAATGCTATTACAACTGGTTATAAGCA..... 159906
362 ThrLysProAlaLeuLysSerGluGluLysThrProIleLysLysProG 378
:::|||||
159907TCCATTAACTCTCAGGATCAA.....CTTGAGAAGACAGG 159941
378 yAspGlyArgLysValThrPheGluProGlySerGlyAspGluAsnG 395
|:::|||||
159942 GCAAAACAAA.....GCATGAGTGATCCACGCCCTGGGTTTCCTGAAC 159985
395 lYThrSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisGln 411
|||
159986 CAACAATAAT..... 159996
412 GlnLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaVa 428
|||
159997 ...ACTCCAAAAGGAATCTTCATTGTCCTCGATGTTATCAAGAGAT 160043
428 lGlyValSerGlnGlyHisHisThrLysAspPheThrArgAlaAlaProA 445
:|:::|||||
160044 GGAAGCCAGCCACCAAGTAACCTCTGCGACTACTCTAGGCTATTGTTGT 160093
445 snPro.....Ala 447
::|
160094 CACCCAGAGATATGAACCAACCTTCAAGCTCTTGTTCAGTATGATCCC 160143
448 LysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyLysTh 464
:::|||||
160144 ACATCAATAGTTCAGCTACATAGCCAGGAACTCCTTATGAATGGAAC 160193
464 rSerProThrAlaGlu...ThrIleLeuLysAsnAsnIleSerSerGlyH 480
|||
160194 ATCTTCTACAGCTGAAGCCACAGGTTTAAAGA.....AGTCTCCTA 160237
480 lSValProHisGlyProLeuThrArgProSerGluGlnLeuAspTyrLeu 496
|||
160238 CTCCCCCTGTTCTCCAGTA..CAACCTTCAAAACAACACTGGAATATTTA 160284
497 SerArgValGlnGlyPheGlnValGluTyrLysAspPheProLysAsnAs 513
:::|||||
160285 GCAAGGATTCAAGGCTTTCAGGTATGAATTAAGAAGC.....AAAAACA 160328
513 nLys 514
:|
160329 AAAA 160332

seq_name: gb_pr4:AL133174

seq_documentation_block:
LOCUS AL133174 113836 bp DNA PRI 14-JUN-2000
DEFINITION Human DNA sequence from clone RP3-470L14 on chromosome 20, complete
sequence.
ACCESSION AL133174
VERSION AL133174.15 GI:8573761
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 113836)
Ramsay, H.
Direct Submission
Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 20, 2000 this sequence version replaced gi:8388434.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

COMMENT

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL, SW:, SWISSPROT, Tr:, TREMBL, Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone conlgs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP3-470L14 is from the library RPCI-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pcypac2
IMPORTANT: This sequence is not the entire insert of clone
RP3-470L14 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP3-470L14 is at 113836 in this
sequence. The true right end of clone RP1-155G6 is at 100 in this
sequence.

FEATURES

source
1. 113836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP3-470L14"
/clone_lib="RPCI-3"
/clone_1lb="RPCI-3"
3061..3077
/note="Weak data"
95583..95627
/note="Single clone region. pUC subcloned to give 12x
coverage. Assembly confirmed by restriction digest data."
BASE COUNT 33974 a 25115 c 24233 g 30514 t
ORIGIN

alignment_scores:
Quality: 753.00 Length: 429
Ratio: 4.328 Gaps: 3
Percent Similarity: 40.559 Percent Identity: 40.326

alignment_block:

US-09-316-048-6 x AL133174/rev ..

Align seg 1/1 to reverse of: AL133174 from: 1 to: 113836

371 lYstThrProIleLysLysProGlyAspGlyArgLysValThrPhePheG 387
|||||
38483 AAGACACCCATTAAGAAACCAAGGAGTGAAGAAAGTAACCTTTTGA 38434
387 uProGlySerGlyAspGluAsnGlyThr..... 396
|||||
38433 ACCTGGCTGTGGGAGTGAATGAGACTAGTGTGATCTTATATATCG 38384

396 396
38383 TCTTCCTCTGAACGTATCTGGATTGTCATGGGCTCCTGAATAACAGCA 38334
396 396
38333 GGAGCCCGACGAGTGAAGTCCCGCATGCAAGTGTCTTATCATTCCTG 38284
396 396
38283 ATTCCCTGCGCTGCCATGCGTCCCAAGGGAGGCCGCGTGGGTTGA 38234
396 396
38233 TGGTGTCAATGCGTTTTCAGGCTGTGGATGGATTCGTGTTACCGT 38184
397 Ser..AsnLysGluAspGluPheArgMetProTyrLeu 408
38183 GTATGTCCCTCCAGGTAAATAAGAGATGAGTTCAGGATGCCCTATCTA 38134
409 SerHisGlnGlnLeuProAlaGlyIleLeuPrometValProGluValAl 425
38133 AGTCATCAGCAGCTGCGTGTGAATCTTCCCATGGTGCCCGAGGTCCG 38084
425 aglnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArg 442
38083 CCAGGCTGTAGGAGTGTACAGGACATCACACCAGAAGATTTTACCAGG 38034
442 laAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 458
38033 CAGCTCCGAATCCGTCCCAAGGCCACGGTAACCTGCATGATAGCCCGAG 37984
459 LeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAs 475
37983 TTGTTGTATGGGGCACCCTGCCACAGCCGAGACCATTTTAAAGAATAA 37934
475 nIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGluG 492
37933 CATCTCTCAGGCCAGCTACCCCATGGACCTCTCAGAGACCCCTGAGC 37884
492 InLeuAspTyrLeuSerArgValGlnGlyPheGln..... 503
37883 AACTGAGCTATCTTCCAGAGTCCAGGATTCAGGTAACGTCTGCGCT 37834
503 503
37833 GAGCTGTGATGGCTGTGCTCAGGACAGTCATCTTTGGCTTCTCTCAGA 37784
503 503
37783 AAATGAACTTTGAACCTTGTACTTTTCTTAATCAAGTCTCTGCCAGG 37734
503 503
37733 GATGGGCTGGGAGCAGTGAAGAAGGTAGAGGTATAGAATTGTA 37684
503 503
37683 AGCATATGTGTAGAGAAATATCTTATCTTGTGTAGTAAGATTTCAA 37634
503 503
37633 CATGCAAGCCCAACCTAAAGCTTCTAACCACTAGCAGTGGTTGTAATT 37584
503 503
37583 TTGATTTCAAAGAACAGACTCATGCTGTACACTTTCTGTCTGTAGGA 37534
503 503
37533 AAGGATATGGCTTCTCCAGTCTTGGGCTTCTGAGACAGACCTAGGTTG 37484

503 503
37483 CTGAAGGTTCTGGGCTGACAGGGCAGGTGTTAGTATGCCACAGTGGG 37434
503 503
37433 CTGAGTTGAAAAAAGAGCTGTGATCTTGTGTGGGGCTTGAGCTGC 37384
503 503
37383 AGCCCTCCAGCCCACTCACTACTGCTTCCCTACCTAGCCTCAGCTCTC 37334
504 ValGluTyrLysAspPheProLysAsnAsnLysAsnG 516
37333 TTTTCCCCCTTAGGTGAATACAAAGACTTCCCAAAACACAGAAGC 37284
516 IupheValSerLeuIleAsnLysSerSerGlnProProLeuIleSerHis 532
37283 AATTGTATCTCTTATCAATTGCTCTCTCAGCCACCTCTGATCAGCCAT 37234
533 GlyIleGlyLysAspValGluSerCysHisAspMet 544
37233 GGATGCGCAAGGATGTGAGTCTGCCATGATATG 37198

seq_name: gb_in3:DROSTAUFEN

seq_documentation_block:

LOCUS DROSTAUFEN 5237 bp mRNA INV 26-APR-1993

DEFINITION D.melanogaster mRNA, complete cds.

ACCESSION M69111
VERSION M69111.1 GI:158505

KEYWORDS
SOURCE Drosophila melanogaster 0-4 hour embryo cDNA to mRNA.

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 5237)
Johnston,D.S., Beuchie,D. and Nusslein-Volhard,C.

AUTHORS Staufen, a gene required to localize maternal RNAs in the
TITLE drosophila egg

JOURNAL Cell 66, 51-53 (1991)

MEDLINE 91300552
FEATURES
source

1..5237
/organism="Drosophila melanogaster"

/db_xref="taxon:7227"
/dev_stage="0-4 hour embryo"

<1..5237
275..3355
/codon_start=1

/protein_id="AA073062.1"

/translation="MQHNVHAARPAPIRAAHSHSHAHMHLPMEQHLGPSLQOO
QPPPPQOPPHRDILHARLNHHHLHADOQQOQOTSSNOAAVAAGAAYHHGINSNS
GSNISSNSNQMKIRQOHQLSSNGLGNQPGPPQAFNPLAGNPALAYNQLPH
PPHMAAHLGYAAPPHYYSQAKPAKYNHGSNANSNSGNSNSNYAPKAILQNT
YRNQKVVPVQEVTPVPEPPVTNNATNSTNSTVIASEVYQEDTSQKPTRQE
PASADHVSTGNDATGALSNEDTSSSGRGKDKTPMCLVNEIARYNKITHQRLTEE
RGAHCKTFYVILMLGDEEYSADGEFKTKAQHLAASKAIEETMKHPKIRSEEGG
PMRTHITPTVELNALAMKLGORTFYLDPTQIPTDSIVPPEFAGHLTAPGPMPO
PPPPAYALRQRLNGFVPIPSQPMHPEFHGQRPFPKPPSRFALPPLGAHYHH
GNGPFPSPVPTPSKITLEVQKQKFGVIGRTLQOAKHDAARALQVLKTOAISASEA
LEDSDMEGDKSPISQVHEIGIKRNMVHFKVLRREGPAHMKNEITACIVGSIVTEGE
GNGKVKSKRAEKMLEVELQKLPLPTPKOTPLKRIKVTPKSGAAREGSVSGTGD
GTMOTGKPERRRRLNPPKDKLIDMDADNPITKLIQLOQTRKEKEPIFELIAKNGNET
ARRREFVMEVSASGTARGTSGNSKLAKRNAQAALFELLEAVOYTPNETOSSSECT
SATMSAVTAPAVEATAEGKVPVATPGMPGILILRONKPAKRRDOIVIVKSNEVS
KEEANKKEVAVAEENSNSANSNGSSSSGDSQATEAASEALNTSGSNTSGVSS
NSSNVGANTDGNHNAESKNTSESSNSTSNTQAGVHMKEQLLYLSKLLDFEVNFSDY
PKGNHNEFLITVLTSTHPPQICHGVGSSEESQNDAAASNALKILSKLGNAMK"

BASE COUNT 1596 a 1291 c 1204 g 1146 t
ORIGIN Chromosome 55A, right arm.

alignment_scores:

Quality: 716.00 Length: 644
Ratio: 2.081 Gaps: 18
Percent Similarity: 53.416 Percent Identity: 32.143

alignment_block:

US-09-316-048-6 x DROSTAU FEN ..

Align seg 1/1 to: DROSTAU FEN from: 1 to: 5237

```
71 ILeThrProthrValGluLeuAsnAlaLeuCysMetLysLeuGlyLysLy 87
|||||
1466 ATTACGCCGACCGGTGAATTGAATGCCCTGGCCATGAAGTTGGACAGCG 1515
87 sPrometTyr.....LysProValAsp..... 94
: |||
1516 CACCTTCTACCTGCTGGATGCCACTCAGATTCTCCACGGAATCCATAG 1565
94 ..... 94
1566 TGCCCGCGAGTTTCCCGCGGCTCATTTACTAAGTGCACACGACGACGCG 1615
95 ...ProTyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArgGlyG 110
|||
1616 ATGCCCGCAACCACTCCACCACTGCTTACGCTCTGCCGACGCGCTTGG 1665
110 Y..... 110
1666 CAATGATTTGTGCCCATTTCCATCGCAGCCGATGCATCCGCATTTCTTTC 1715
111 .....AlaTyrProProArgTyrPheTyrProPhePro 121
|||
1716 ATGACCTGGTCAAGCAGCGTTTCTCCGAAGTTTCCATCAGCGCTTTGCG 1765
122 ValProProLeuLeu..... 126
|||
1766 CTGCCACCGCTCTCGGAGCTCATGTACATGACCGAATGGCCCATTT 1815
127 .....TyrGlnValGluLeuSerValGlyGly 136
|||
1816 TCCATCCGTGCTACGCCACCCAGCAAAATCACCTGTTGTGGCAAGC 1865
136 lnglnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspAla 152
|||
1866 AGAAGTTTGTGGCATCGGACGACACTGCAGCAAGCAAAACATGACGCT 1915
153 AlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeu.....Pr 166
|||
1916 GCGCGAGAGCGCTGCAGGTTCTCAAGACACAGGCCATATCTGCCCTTGA 1965
166 ogluArgLeuGluValAsnGlyArgGluSerGluGluAsnLeuAsnL 183
|||
1966 GGAGGCTCTCGAG.....GATTCGATGGATGAGGGGTGACAAGA 2003
183 ysserGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuPro 199
|||
2004 AGTGGCCCATCTCGCAGGTTTCACGAGATCGGAATTAAAGCGCAACATGACC 2053
200 ValAsnPheGluValAlaArgGluSerGlyProProHisMetLysAsnPh 216
|||
2054 GTTCATTTTAAGGTGTTCGCGAGAGGCGCGGCACATGAGAACAATT 2103
216 eValThrLysValSerValGlyGluPheValGlyGluGlyLys 233
|||
2104 TATAACAGCCCTGCATTTGTGGATCTATTTACCGAAGAGAGAAGAATG 2153
233 erLysLysIleSerLysLysAsnAlaIleAlaValLeuGluGluLeu 249
|||
2154 GCAAAAAAGTGTCCAAAGAACGCGGCAAAAGATGCTCTGAGATTG 2203
250 LysLysLeuProLeu...ProAlaValGluArgValLysProArgI 265
```

```
2204 CAAAGCTGCCCGCCCTTACGCCACCAAGACAGCCCATTTGAACGGAT 2253
265 eLysLysLysThr..... 269
|||
2254 TAAGTTAAGACACCAAGCAAAAGTGTGCAGCTGCCCGAGAGGTTCCG 2303
270 .....LysProIleValLys 274
|||
2304 TTGTGTGGGCGACTGATGTACTATGCAAAAGCGGAAGCCGAGAGACGT 2353
275 ProGlnThrSerPro.....GluTyrGlyGlnGlyI 285
|||
2354 AAACGCCCTCAATCCGCCGAAGGACAAAGCTGATAGACATGAGACGATGCCGA 2403
285 easnProIleSerArgLeuAlaGlnIleGlnAlaLysLysGlyLysG 302
|||
2404 TAATCCCATTACAAGCTTAATTCAACTGCAGCAGACGCGTAAGCAAAAG 2453
302 luProGluTyrThrLeuLeuThrGluArgGly....LeuProArgArg 316
|||
2454 AACCCATTTTCGAGCTGATAGCCAAAACGGCAATGAACCCGCTCGACGA 2503
317 ArgGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyTh 333
|||
2504 CGGGAATTGTTATGAGGCTCTCCGCCAGCGGAGACACAGCCCGTGGAAAC 2553
333 rGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuG 350
|||
2554 AGGCACAGCAGCAAGAAATTGGCAAAAGCAAAATGCTGCACAGAGCTCTATTG 2603
350 luIleLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeu 366
|||
2604 AACTGCTG.....GAAGCTGTCAAGTGCACACCAACCAAGCAACACAG 2647
367 LysSerGluGlu..... 370
|||
2648 TCATCGAGGAATGCTGCACCTTCGGCAACAATGTCCGCCGTTACAGCGCC 2697
371 .....LysThrProI 374
|||
2698 CGCAGTGGAGGCAACTGCCGAAGGCAAAAGTGCCCATGTAGCAACGCCAG 2747
374 le.....LysLysProGly 378
|||
2748 TGGGTCCCATGCCAGGCATTTTAATACTGCCGCCAGAACAAAGCAAGCA 2797
379 AspGlyArg...LysValThrPhePheGluProGlySerGlyAspGluAs 394
|||
2798 AAGAAGAGGATCAAAATTGTAATTGTCAAGTCTAACGTGGAATCCAAAGA 2847
394 nGlyThrSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisG 411
|||
2848 GGAGAGGCGCAATTAAGGAGTGGCAGTTGCAGCTGAGAGAAATAGCAACA 2897
411 lnglnLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAla 427
|||
2898 ACAGTGCAAACTCTGGC.....GACAGCAGCAACACAGCAGC 2932
428 ValGlyValSerGlnGlyHisHisThrLysAspPheThrArgAlaAlaPr 444
|||
2933 AGCGGCGATTCCGAAGCC.....ACAGAGGCGGCTAG 2964
444 oAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuT 461
|||
2965 TGAGAGTGCATGAATACCTCAACCGCAGCAATACG..... 3001
461 yrglyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnIleSer 477
|||
3002 ..AGTGTGTGAGTAGCAACAGCAGCAATGTGAGGCCAACACGATGTGT 3049
478 SerGlyHisValProHisGlyProLeuThrArgProser..... 490
|||
```

```
3050 AACACCATGCCGAGAGTAATAAACACACGAGAGTAGCAGCAACAGTAC 3099
491 .....GluGlnLeuAspTyrL 496
3100 GAGTAACACACAAAGCGCTGGAGTGCACATGAAGAGCAGCTCTTGTAAC 3149
496 euSerArgValGlnGlyPheGlnValGluTyrLysAspPheProLysAsn 512
3150 TTAGTAAACTGCTGACTTTGAGGTCAACTTCTCGACTATCCGAAAGGC 3199
513 AsnLysAsnGluPheValSerLeuIleAsnCysSerSerGlnProProLe 529
3200 AATCACAACGAGTTCCTGACCATCGTGACATGTCCACACATCCGCCGCA 3249
529 uIleSerHisGlyIleGlyLysAspValGluSerCysHisAspMetAla 546
3250 GATTGGCCATGGCGTTGGCAAGAGCTCCGAGGAGTCGCAGAATGATGCTG 3299
546 IaleuAsnIleLeuLysLeuLeuSerGlnLeu 556
3300 CAAGCAATGCCCTTGAAAAATCCTCAGCAAGCTG 3331
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2001, 16:02:36 ; Search time 701.87 Seconds
(without alignments)
1681.696 Million cell updates/sec

Title: US-09-316-048-1

Perfect score: 3142

Sequence: 1 acttcctgcgcggctgcggg.....taataactaaaaaaaaa 3142

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36:*

- 1: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3138.8	99.9	3142	21	Z93324	Human stauflen CDNA
2	3053.8	97.2	3217	21	Z93323	Human stauflen CDNA
3	2971.4	94.6	3190	20	X90786	DNA encoding human
4	2896	92.2	3506	21	Z93322	Human stauflen CDNA
5	2890.2	92.0	3260	21	Z93327	Human stauflen CDNA
6	1356.2	43.2	2859	21	Z93328	Mouse stauflen CDNA
7	1107.2	35.2	1127	20	X39991	Prostate cancer as
8	798.8	25.4	946	20	X40002	Prostate cancer as
9	755.6	24.0	844	20	X40000	Prostate cancer as
10	753.2	24.0	773	20	X39992	Prostate cancer as
11	551	17.5	825	20	X40001	Prostate cancer as
12	501.4	16.0	794	20	X39993	Prostate cancer as

C	13	433.8	13.8	660	21	Z79981	Human colon cancer
	14	412.2	13.1	727	21	Z80269	Human colon cancer
C	15	327.8	10.4	708	21	Z80703	Human colon cancer
	16	286.6	9.1	305	16	T20075	Human gene signatu
	17	130.8	4.2	300	20	Z17757	Human gene express
	18	130.8	4.2	300	20	Z14226	Human gene express
	19	126.2	4.0	769	20	Z17411	Human gene express
	20	87.8	2.8	767	20	Z16120	Human gene express
C	21	55	1.8	1000	21	A02484	Human gene express
	22	52	1.7	1337	20	Z17263	Human gene express
	23	52	1.7	12001	16	O76213	HSV L/ST region.
	24	51.8	1.6	2188	20	Z77506	Human ovarian tumo
	25	51.6	1.6	936	20	X56375	Human DNA-dependen
	26	51.2	1.6	53526	19	T94101	Human PKD1 gene.
	27	51.2	1.6	53577	17	T18551	Human polycystic k
	28	51.2	1.6	53577	19	T94108	Human polycystic k
	29	50.4	1.6	10732	21	A10594	Human encoding a su
	30	48.8	1.6	867	21	Z98056	Gene encoding pro
C	31	48.6	1.5	15899	19	V62164	Human secreted pro
	32	48.6	1.5	16812	19	V62175	HSV-2 strain SB5 C
C	33	48.4	1.5	38186	20	Z32028	Human METH1 relate
C	34	48	1.5	8670	20	Z32059	Human METH2 relate
	35	47.6	1.5	3198	20	X02974	Human IL-1ra BAC c
	36	47	1.5	1235	20	Z16147	Human gene express
	37	47	1.5	23142	21	A35158	Human gene express
	38	47	1.5	30417	21	A35160	Human adenosine re
	39	45.4	1.4	80240	20	V83940	Human adenosine re
	40	45.4	1.4	80595	20	V83939	NC-contig derived
C	41	45.2	1.4	1446	20	Z16805	HC-contig derived
C	42	45	1.4	1523	20	Z17507	Human gene express
C	43	45	1.4	1523	20	Z17479	Human gene express
	44	45	1.4	4897	11	Q03259	Human gene express
	45	44.8	1.4	1725	12	Q11243	Pseudorabies virus PRV 28k gene. Pse

ALIGNMENTS

RESULT 1	
ID Z93324	standard; cDNA; 3142 BP.
XX AC Z93324;	
XX DT 04-JUL-2000	(first entry)
XX DE Human stauflen CDNA.	
XX KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW acquired immune deficiency syndrome; protease; human; ss.	
XX OS Homo sapiens.	
XX FH Key	Location/Qualifiers
FT CDS	288..1193
FT FT	/*tag= a
FT FT	/product= Stauflen protein
XX PN CA2238656-A1.	
XX PD 22-NOV-1999.	
XX PF 22-MAY-1998;	98CA-2238656.
XX PR 22-MAY-1998;	98CA-2238656.
XX PA (UYMO-) UNIV MONTREAL.	
XX PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;	
XX PI Moulard AJ;	
XX DR WPI; 2000-246924/22.	

DR P-PSDB; Y83108.

XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)

PS Claim 4; Figure 1b; 96pp; English.

XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC Z93322.

XX
SQ Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 99.9%; Score 3138.8; DB 21; Length 3142;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3140; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 acttcctgcgggctgcgggcctgagcgctcttcagcggttgcgcgcgctgcgcgt 60
DB 1 acttcctgcgggctgcgggcctgagcgctcttcagcggttgcgcgcgctgcgcgt 60
QY 61 ctctctgcgctccgcttctcttgaacgcctcccccccgcgcgcgcgcgctc 120
DB 61 ctctctgcgctccgcttctcttgaacgcctcccccccgcgcgcgcgcgctc 120
QY 121 ctccagcgcaactcgcctcttccctccttcgtcccttctccttctccttcttcc 180
DB 121 ctccagcgcaactcgcctcttccctccttcgtcccttctccttctccttcttcc 180
QY 181 ctccctccctcctcctgcgcgcacgcgccagagacgcgcggggagagctcgagc 240
DB 181 ctccctccctcctcctgcgcgcacgcgccagagacgcgcggggagagctcgagc 240
QY 241 agcagccagaaagcataaccctactgtagaactaatgcaactgtgcatgaacttgaa 300
DB 241 agcagccagaaagcataaccctactgtagaactaatgcaactgtgcatgaacttgaa 300
QY 301 aaaaaccaatgtataagcctgttgaaccttactctcgatgcagtcacactatactaca 360
DB 301 aaaaaccaatgtataagcctgttgaaccttactctcgatgcagtcacactatactaca 360
QY 361 acatgagagaggtgtctatccccgaggtactttaccatttccagttccacctttac 420
DB 361 acatgagagaggtgtctatccccgaggtactttaccatttccagttccacctttac 420
QY 421 ttatcaagtgaacttctgtgtgagagacagcaatttaatgcaaaagaaagacaagac 480
DB 421 ttatcaagtgaacttctgtgtgagagacagcaatttaatgcaaaagaaagacaagac 480
QY 481 agctgcgaaacacagatgtctgtgccaaagcgttgagatcctgcagaatgagccctgc 540
DB 481 agctgcgaaacacagatgtctgtgccaaagcgttgagatcctgcagaatgagccctgc 540
QY 541 cagaagagctgtgagtgtaatggaagagaatccgaaagaaaaatctcaataaactgaaa 600
DB 541 cagaagagctgtgagtgtaatggaagagaatccgaaagaaaaatctcaataaactgaaa 600

QY 601 taagtcaagtgtttagatgtgcacttaaacggaacttgccctgtgaatttcgaggtgccc 660
DB 601 taagtcaagtgtttagatgtgcacttaaacggaacttgccctgtgaatttcgaggtgccc 660
QY 661 gggagagtgccccaccaccatgaagaacttgtgaccaaggttcggttgggagtttg 720
DB 661 gggagagtgccccaccaccatgaagaacttgtgaccaaggttcggttgggagtttg 720
QY 721 tggggagaagtgaagggaagaaagcaagaagatttcaaagaaaaatgccgcatagtcttc 780
DB 721 tggggagaagtgaagggaagaaagcaagaagatttcaaagaaaaatgccgcatagtcttc 780
QY 781 ttgagagctggaagaagttaaccgccccctgctgtcagttgaaagtaagcctagaatca 840
DB 781 ttgagagctggaagaagttaaccgccccctgctgtcagttgaaagtaagcctagaatca 840
QY 841 aaaaagaaaaaaaccatagtcacagccacagacaagcccaagatatgcccagggatca 900
DB 841 aaaaagaaaaaaaccatagtcacagccacagacaagcccaagatatgcccagggatca 900
QY 901 atccgattagccgacttggcccagatccagcagcgaagcaaaaaagagaagagccagatca 960
DB 901 atccgattagccgacttggcccagatccagcagcgaagcaaaaaagagaagagccagatca 960
QY 961 cgtctcacaagagcagagcctcccgcgccgcagggagttgtgatgcagttgaaggttg 1020
DB 961 cgtctcacaagagcagagcctcccgcgccgcagggagttgtgatgcagttgaaggttg 1020
QY 1021 gaaaccacactgcagaaggaacgggcaccaacaagaagtgtgccaaagcgaatgcagccg 1080
DB 1021 gaaaccacactgcagaaggaacgggcaccaacaagaagtgtgccaaagcgaatgcagccg 1080
QY 1081 agaacatgtgtgagatccttgtttcaaaagtcctgcgcgcgcagccaccacaacccgcac 1140
DB 1081 agaacatgtgtgagatccttgtttcaaaagtcctgcgcgcgcagccaccacaacccgcac 1140
QY 1141 tcaagtcagaggaagaaagacacccaataaagaaaccagggatggaagaaagttaaccttt 1200
DB 1141 tcaagtcagaggaagaaagacacccaataaagaaaccagggatggaagaaagttaaccttt 1200
QY 1201 ttgaacctgtcttggggatgtaaaatgagactagtaataaagagatgagttcagatgc 1260
DB 1201 ttgaacctgtcttggggatgtaaaatgagactagtaataaagagatgagttcagatgc 1260
QY 1261 ctatctaaagtcacagcagctgtcctgtgtaattcttcccatgtgtgccgagtcgccc 1320
DB 1261 ctatctaaagtcacagcagctgtcctgtgtaattcttcccatgtgtgccgagtcgccc 1320
QY 1321 aggtgtgagagttagtcaagagcatcacaccaagaatttaccagggcagctcogaatc 1380
DB 1321 aggtgtgagagttagtcaagagcatcacaccaagaatttaccagggcagctcogaatc 1380
QY 1381 ctgcgaagccaacggttaactgcacatagaccgagagttgtgtatgtggygcacctgc 1440
DB 1381 ctgcgaagccaacggttaactgcacatagaccgagagttgtgtatgtggygcacctgc 1440
QY 1441 ccacagccgagaccattttaagaataacatctcttcagggccaagtaaccatgagacctc 1500
DB 1441 ccacagccgagaccattttaagaataacatctcttcagggccaagtaaccatgagacctc 1500
QY 1501 tcaagagacctctgagcaactgactatcttccagagttccagggatccaggttgaat 1560
DB 1501 tcaagagacctctgagcaactgactatcttccagagttccagggatccaggttgaat 1560
QY 1561 acaaaagcttccccaaaaaacaagaagaacgaattgtatctcttaacaattgctcctc 1620
DB 1561 acaaaagcttccccaaaaaacaagaagaacgaattgtatctcttaacaattgctcctc 1620
QY 1621 agccacctgtatcagccatgtgtatcggcaagatgtgagtcctgtccatgatgtgctg 1680
DB 1621 agccacctgtatcagccatgtgtatcggcaagatgtgagtcctgtccatgatgtgctg 1680
QY 1681 cgctgaacatcttaagtgtgtctgtgagttgacccaacaaagtacagagatgccaagaa 1740

Db 1681 ||||| cgtgaacatcttaagtgtgtctgtgagtgagccaacaagtlacagagatgtccaagaa 1740
QY 1741 caggaacgagaccaatgtctgtgtgtggaggtgtctgaaccttctctggcatgaacct 1800
Db 1741 ||||| caggaacgagaccaatgtctgtgtgtggaggtgtctgaaccttctctggcatgaacct 1800
QY 1801 tataaaatcccaacataataactgaaaaataactgtaaccttttgaaaaatttggaaattct 1860
Db 1801 ||||| tataaaatcccaacataataactgaaaaataactgtaaccttttgaaaaatttggaaattct 1860
QY 1861 gataacctcagtggtgcgagagacacagtggtgtaaaagtgtgtggcagcagcaggaaga 1920
Db 1861 gataacctcagtggtgcgagagacacagtggtgtaaaagtgtgtggcagcagcaggaaga 1920
QY 1921 caacagaacaacaagagagcggtgtgtgcggtgactgtgtgtgggtttgtgtgatg 1980
Db 1921 ||||| caacagaacaacaagagagcggtgtgtgcggtgactgtgtgtgggtttgtgtgatg 1980
QY 1981 gccactcgtgacctgtgcgtccctacgcgaatagcagctgcctgtgtgggaaagaagcgctg 2040
Db 1981 gccactcgtgacctgtgcgtccctacgcgaatagcagctgcctgtgtgggaaagaagcgctg 2040
QY 2041 cccagccagctgtgttcccccgggacaccagcagatccacaccctgggcacctcgtgtt 2100
Db 2041 cccagccagctgtgttcccccgggacaccagcagatccacaccctgggcacctcgtgtt 2100
QY 2101 ggtcttttttccctgtgtgaagaagaacggcagcagcccttctcaagctgtgtca 2160
Db 2101 ggtcttttttccctgtgtgaagaagaacggcagcagcccttctcaagctgtgtca 2160
QY 2161 ctcaagacacatgtggagacaaaccttgagacagccatgccaagagagcgcttgaccggccc 2220
Db 2161 ctcaagacacatgtggagacaaaccttgagacagccatgccaagagagcgcttgaccggccc 2220
QY 2281 agtgtgccacggccccaccacctctctgcagtlaccacacacatccacctgttctcttc 2340
Db 2281 agtgtgccacggccccaccacctctctgcagtlaccacacacatccacctgttctcttc 2340
QY 2341 caacagtgatctgtatctttagtttcatatttcttlttgatgataagacatatataa 2400
Db 2341 ||||| caacagtgatctgtatctttagtttcatatttcttlttgatgataagacatatataa 2400
QY 2401 aatttcaatttgagaatttctcaattgtatcttagttaaatagcacagtttgaaaacttgt 2460
Db 2401 aatttcaatttgagaatttctcaattgtatcttagttaaatagcacagtttgaaaacttgt 2460
QY 2461 ctgagactgactttatcaataatctaacgcgacaaagatcatatcatgtgtatgtgtta 2520
Db 2461 ctgagactgactttatcaataatctaacgcgacaaagatcatatcatgtgtatgtgtta 2520
QY 2521 gacatttttatttcatgtactaaaccagagacagtttcagtgatgcaaatgtgtgccctc 2580
Db 2521 gacatttttatttcatgtactaaaccagagacagtttcagtgatgcaaatgtgtgccctc 2580
QY 2581 tggttcagctgaaacagtcctgtgactttcaaaaaccttgaataagtcctccacagttgta 2640
Db 2581 tggttcagctgaaacagtcctgtgactttcaaaaaccttgaataagtcctccacagttgta 2640
QY 2641 taaattggacaatttagaattttaaactttagatgatacatttggttccattttatttc 2700
Db 2641 taaattggacaatttagaattttaaactttagatgatacatttggttccattttatttc 2700
QY 2701 attttattttgttaatgcaaacagagactaaatgaacttgatctctgttttaaat 2760
Db 2701 attttattttgttaatgcaaacagagactaaatgaacttgatctctgttttaaat 2760
QY 2761 tattaaaaaacatgtgtatctatacatatgtgctctgagacttagcttcaactact 2820
Db 2761 ||||| tattaaaaaacatgtgtatctatacatatgtgctctgagacttagcttcaactact 2820

Db 2761 tattaaaaaacatgtgtatctatacatatgtgctctttaggacttagcttcaactact 2820
QY 2821 acagatatgatctccatgtagtlccatataaaacctgcagagtgtatttccagagtgtcg 2880
Db 2821 ||||| acagatatgatctccatgtagtlccatataaaacctgcagagtgtatttccagagtgtcg 2880
QY 2881 atactgttaattacatctccattagggctgaaaagaatgacctagctgttctgtatacagc 2940
Db 2881 atactgttaattacatctccattagggctgaaaagaatgacctagctgttctgtatacagc 2940
QY 2941 tgtgtgtcttttgatgtgtgttactgttacacagaagtgtgtgcactgaggtctgtcg 3000
Db 2941 ||||| tgtgtgtcttttgatgtgtgttactgttacacagaagtgtgtgcactgaggtctgtcg 3000
QY 3001 tgttcggtatgaaaaacctggtagccctgcgagtttaagtactgtctccatcattgttta 3060
Db 3001 tgttcggtatgaaaaacctggtagccctgcgagtttaagtactgtctccatcattgttta 3060
QY 3061 cgctggaatttctcccatggaatgtaagtaaaacttaagtgttgcatacaataaat 3120
Db 3061 ||||| cgctggaatttctcccatggaatgtaagtaaaacttaagtgttgcatacaataaat 3120
QY 3121 ggtactactaaaaaa 3142
Db 3121 ggtactactaaaaaa 3142

RESULT 2
Z93323
ID Z93323 standard; cDNA; 3217 BP.
XX
AC Z93323;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 363..1853
FT /*tag= a
FT /product= Stauflen
XX
CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desroselliers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96pp; English.
XX
CC Stauflen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds

RNA-binding domains. Chimeric proteins comprising portions of stauflen with proteins with RNase or protease activity can be used to treat virus infections. The RNase or protease activity of the fusion protein prevents proper maturation of the virus. Mammalian stauflen proteins seem to recognise double stranded RNA structure rather than any sequence specific position. 2 bacterially-expressed fusion proteins used in an RNA-binding assay (his/hsp (human SP) and MBP/mSP (murine SP)) both proteins strongly bound double stranded RNA. Both fusion proteins also directly bound labeled ds RNAs and RNA/DNA hybrids. Stauflen proteins may therefore be useful in the treatment of RNA virus infections, especially those caused by retroviruses, in particular human immunodeficiency virus. This sequence is an alternatively spliced sequence to the one given in Z93322.

Sequence 3217 BP; 889 A; 796 C; 742 G; 790 T; 0 other;

Query Match	97.28;	Score 3053.8;	DB 21;	Length 3217;
Best Local Similarity	97.68;	Pred. No. 0;		
Matches 3140; Conservative	0;	Mismatches	2;	Indels 75; Gaps 1;

```

Qy      1 acttcctgcgcggcctgcgcgcctcctcagcgtttgcgcgcgcgcctgcgcgt 60
        |||||
Db      1 acttcctgcgcgcgcctgcgcgcctcctcagcgtttgcgcgcgcgcctgcgcgt 60

```

Qy 61 ctctctcgctcccgcttctcttgaacgcctcccccccgcccgcgcgcccgctc 120
|||||
db 61 ctctctcgctcccgcttctcttgaacgcctcccccccgcccgcgcgcccgctc 120

Qy	Db
121 ctccacgacacacgcgcctctcccccctctgcctctctccctctccctttttccct	121 ctccacgacacacgcgcctctcccccctctgcctctctccctctccctttttccct
121 ctccacgacacacgcgcctctcccccctctgcctctctccctctccctttttccct	121 ctccacgacacacgcgcctctcccccctctgcctctctccctctccctttttccct

```

Oy      181 ctccctccctctctgcgcgcacgcgccagagacgcgcgcggggagcagctcgcgagc 240
          |||||
Db      181 ctccctccctctctctgcgcgcacgcgccagagacgcgcgcgggggacgagtcgcgagc 240

```

Oy	241	agcagcc-----	247
Db	241	agcagccaggttcaatcaacttaacctcagaactgaacaagacacattgttc	300

```

Qy      248 -----agaaagcataaccctactgtagaactaatgcactgt 285
          |||
Db      301 tgyaacgcctctttttaaanaagaaagcataaccctactgtagaactaatgcactgt 360

```

QY 286 gcatgaaacttgaaaaaaacccaatgtataagcctgttgacccttaactctcgatgcagt 345
|||||
Db 361 gcatgaaacttgaaaaaaacccaatgtataagcctgttgacccttaactctcgatgcagt 420

[illegible]

```

QY 406 cagttcaccttaattatcaagtggaattctgtggagagacagcaatttaatgca 465
    |||||
Db 481 cagttcaccttaattatcaagtggaattctgtggagagacagcaatttaatgca 540

```

Qy 466 aaagaaagacaagacagctctgcgaacacgatgtctctgccaagcgttagagatccctgc 525
|||||
Db 541 aaagaaagacaagacagctctgcgaacacgatgtctctgccaagcgttagagatccctgc 600

QY 526 agaatgagcccccgcagagagcctgtaggttgaaatggaagagaaatccgaagaagaataatc 585
|||||
Db 601 agaatgagcccccgcagagagcctgtaggttgaaatggaagagaaatccgaagaagaataatc 660

Qy	586 tcaataaatctgaaataagtcgaagtgtttggaggattgcacttaacgyaaccttgcccttgya 645
Db	661 tcataatatcrgaaaatagtcgaagtgtttggaggattgcacttaacgyaaccttgcccttgya 720

QY 646 attctgaggtgcccgggagagtggtgcccacacatgtaagaacttgtgaccaagttt 705
|||||
Db 721 attctgaggtgcccgggagagtggtgcccacacatgtaagaacttgtgaccaagttt 780
|||||

QY	706	cggcttgaggagttcgttggggaaggtgaagggaataattcaagaataatg	765
Db	781	cggcttgaggagttcgttggggaaggtgaagggaataattcaagaataatg	840

OY 766 ccgcacatagctgttctcttgagaagctgaagaagttaaccgcccctgctcgtcagtttgaacgag 825
|||||
Db 841 ccgcacatagctgttctcttgagaagctgaagaagttaaccgcccctgctcgtcagttgaacgag 900

```
QY      826 taagcctagaatcaaaaagaaacaaccatagtcaagccacagacaagcccgaat 885
          |||||
Db      901 taaagcctagaatcaaaaagaaacaaccatagtcaagccacagacaagcccgaat 960
```

Qy 886 atgcccaggggatcaatccgatctagccgactggtccagatccagcagaagcaaaaaagagaga 945
|||||
961 atgcccaggggatcaatccgatctagccgactggtccagatccagcagaagcaaaaaagagaga 1020

QY 946 agtagccagagtacacgctccctcaacagagcgagcgctcccgcccgacaggagttgtga 1005
|||||
|||
Db 1021 agtagccagagtacacgctccctcaacagagcgagcgctcccgcccgacaggagttgtga 1080

[illegible]

QY	1066 agcgcaatgcagccgagaacatacttgagatccttggttccaagtcccgcagcggcagc 	1125
Db	1141 agcgcaatgcagccgagaacatgttgagatccttggttccaagtcccgcagcggcagc 	1200

QY	1126	ccaccaaacccgcactcaagtcagaaggaaacaccatgaataaaccaagggtgtaa	1185
Db	1201	ccaccaaacccgcactcaagtcagaaggaaacaccatgaataaaccaagggtgtaa	1260

QY 1186 gaaaagtaaccttttttgaaacctggtctgggatgaaaaatgaggactagtataaagaagg 1245
|||||
|||
Db 1261 gaaaagtaaccttttttgaaacctggtctgggatgaaaaatgaggactagtataaagaagg 1320

[illegible]

Dy 1306 tgcctcaggtcgcccaagtctgtagtgattatgaagaacaccaccaaaagatttaca 1365

Db 1381 tgcccaggctgcccaagctgttagygattagtaagaacatacacccaagaatttacca 1440

Qy 1366 gggcagctccgaatcctctgccaagggccacgtaactgcatgatagcccgagattgttgc 1425
|||||
Db 1441 gggcagctccgaatcctctgccaagggccacgtaactgcatgatagcccgagattgttgc 1500

Qy 1426 atgggggcacctgcgccacagccgagaccattttaagaatacatctcttcagggcacg 1485
|||||
Db 1501 atgggggcacctgcgccacagccgagaccattttaagaatacatctcttcagggcacg 1560

[illegible]

Qy 1546 gattccaggttgatatacaagagcttccccaaaaacaacaagaagaaattgtatctcta 1603
|||||
Db 1621 gattccaggttgatatacaagagcttccccaaaaacaacaagaagaaattgtatctcta 1680

Qy 1606 tcaattgtctctctacagccacccctgtatcagccatgtgtatcgcgaagatgttgagctcc 1600
 |||||
 Db 1681 tcaattgtctctctacagccacccctgtatcagccatgtgtatcgcgaagatgttgagctcc 1740

Dy	1666	gccaatgataatgcctgcgtcgaacacattaaagtctgtctcagttggaccaacaaga	1722
Db	1741	gccaatgatatgctgcgtcgtaaacatttaaagttgtctgtcctgagtttgaccacaaga	1800

QY 1726 cagagatgccaagaacaggaacgacccaatgtctgtgtggaaggtgtcctgaacctttc 1765
|||||
Db 1801 cagagatgccaagaacaggaacgacccaatgtctgtgtgtggaaggtgtcctgaacctttc 1860

QY 1786 ctggccatgaaacctataaatacccaacatatatactgaaaaataactgacttga 1845
|||||
Db 1861 ctggccatgaaacctataaatacccaacatatatactgaaaaataactgacttga 1920
QY 1846 aaatttggaattctctgatacctccagtgggccgagagacacggtggttaaagatgtgg 1905
|||||
Db 1921 aaatttggaattctctgatacctccagtgggccgagagacacggtggttaaagatgtgg 1980
QY 1906 cagcagcagggagagacaacagaaacacagagagcggtgtgcccggctgtagctgtgtg 1965
|||||
Db 1981 cagcagcagggagagacaacagaaacacagagagcggtgtgcccggctgtagctgtgtg 2040
QY 1966 ggggttgtgtgataggccaactcgtgtagcctggcggtccctacgcaatagcagctgctgt 2025
|||||
Db 2041 ggggttgtgtgataggccaactcgtgtagcctggcggtccctacgcaatagcagctgctgt 2100
QY 2026 gggggaagaagggtgcccagccagcagctgttctcccgggacacccagcagatccacacctg 2085
|||||
Db 2101 gggggaagaagggtgcccagccagcagctgttctcccgggacacccagcagatccacacctg 2160
QY 2086 ggcacctccgtgtgtgtgtcttttttccctgtgtgaaagaagaacggcagcagccct 2145
|||||
Db 2161 ggcacctccgtgtgtgtgtcttttttccctgtgtgaaagaagaacggcagcagccct 2220
QY 2146 tctcaagctgtgctcaactcagacacatctgggacaacacctgagacagcatgccaagagag 2205
|||||
Db 2221 tctcaagctgtgctcaactcagacacatctgggacaacacctgagacagcatgccaagagag 2280
QY 2206 gcctttgaccggcccagagctaaagaacacagagaaatcaaatgcttccctactcagcg 2265
|||||
Db 2281 gcctttgaccggcccagagctaaagaacacagagagaaatcaaatgcttccctactcagcg 2340
QY 2266 tgacccaactttctctagtgtgcccacggcccacccacctcctgcaagtaaccacacacacac 2325
|||||
Db 2341 tgacccaactttctctagtgtgcccacggcccacccacctcctgcaagtaaccacacacacac 2400
QY 2326 cactgcttctctctccaacagtgatctgtatctttagtttcatatttctttagattga 2385
|||||
Db 2401 cactgcttctctctccaacagtgatctgtatctttagtttcatatttctttagattga 2460
QY 2386 tatgacacataataaatttcatatttgagaattctcaattgtatctagttaaatagcac 2445
|||||
Db 2461 tatgacacataataaatttcatatttgagaattctcaattgtatctagttaaatagcac 2520
QY 2446 agtttgaaactgtctgagactgacttcatcaataatctaacccgacaagaatcatatcc 2505
|||||
Db 2521 agtttgaaactgtctgagactgacttcatcaataatctaacccgacaagaatcatatcc 2580
QY 2506 atgtgtatgtgttagacatttttatttcatgtactaacccagagacagtttcagtgatgc 2565
|||||
Db 2581 atgtgtatgtgttagacatttttatttcatgtactaacccagagacagtttcagtgatgc 2640
QY 2566 aaattgtgtccctcgtgttcagctgaaaaagctcctgagcttcaaaaaacctggaataag 2625
|||||
Db 2641 aaattgtgtccctcgtgttcagctgaaaaagctcctgagcttcaaaaaacctggaataag 2700
QY 2626 tctccacacagttgtataaattggaacaatttaggaattttaaactttagatgatttg 2685
|||||
Db 2701 tctccacacagttgtataaattggaacaatttaggaattttaaactttagatgatttg 2760
QY 2686 ttccattttatttcaattttattttgtttaaagcaaacagagacttaaatgaactttgat 2745
|||||
Db 2761 ttccattttatttcaattttattttgtttaaagcaaacagagacttaaatgaactttgat 2820
QY 2746 ctctgttttaagaattataaanaaacatgtgtatatctatacatatgctcttgagagact 2805
|||||
Db 2821 ctctgttttaagaattataaanaaacatgtgtatatctatacatatgctcttgagagact 2880
QY 2806 agcttccactacacagagatagatctccatgttagtccatataaacctgcagagtgat 2865
|||||
Db 2881 agcttccactacacagagatagatctccatgttagtccatataaacctgcagagtgat 2940
QY 2866 ttccagagtgctcgatactgttaattacatctccattagggctgaaaaaatgacacctac 2925

Db 2941 ttccagagtgctcgcatactgttaattacatctccattagggctgaaaaaatgacacctac 3000
QY 2926 gttctgtatacagcgtgtgtgtcttttgatgtgtgttactgttacacagaagtgtgtga 2985
|||||
Db 3001 gttctgtatacagcgtgtgtgtcttttgatgtgtgttactgttacacagaagtgtgtga 3060
QY 2986 ctgaggtctcgtgtgtgtccgtatggaanaacctgtgacctgcgagtttaagtactgct 3045
|||||
Db 3061 ctgaggtctcgtgtgtgtccgtatggaanaacctgtgacctgcgagtttaagtactgct 3120
QY 3046 tccattcatgtttagcctggaatttcttcccatggaatgtgaagtaaaacttaagtgt 3105
|||||
Db 3121 tccattcatgtttagcctggaatttcttcccatggaatgtgaagtaaaacttaagtgt 3180
QY 3106 ttgtcatcaataaatggttaataactaaaaa 3142
|||||
Db 3181 ttgtcatcaataaatggttaataactaaaaa 3217

RESULT 3
X90786
ID X90786 standard; DNA; 3190 BP.
XX
AC X90786;
XX
DT 13-JAN-2000 (first entry)
XX
DE DNA encoding human staufen protein.
XX
KW Human staufen protein; hStau; protein lysate; lung; kidney; testis;
KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;
KW telomerase assembly; maturation; transport; regulation; interaction;
KW inhibitory agent; antibody; immunoassay; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 345..1835
FT /*tag= a
FT /product= "Human staufen protein"
FT /note= "Binds to telomerase RNA"
FT poly_a_signal 3172..3177
FT /*tag= b
XX
PN WO9951255-A1.
XX
PD 14-OCT-1999.
XX
PF 06-APR-1999; 99WO-US07533.
XX
PR 06-APR-1998; 98US-0080783.
XX
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Greider CW, Le S;
XX
DR WPI; 1999-620168/53.
DR P-PSDB; Y26335.
XX
PT Human staufen polypeptide useful in methods for identifying telomerase
PT inhibitors -
PS Clalm 13; Page 41-42; 50pp; English.
XX
CC The present sequence is an isolated DNA encoding human staufen (hStau)
CC protein. The hStau protein was detected in protein lysates from lung,
CC kidney, testis and ovary, but not in brain and heart. It binds
CC specifically to human telomerase RNA and can complex with the
CC telomerase catalytic subunit. The hStau protein may play a role in
CC telomerase assembly, maturation, transport and regulation. Interaction
CC between hStau and telomerase can be inhibited by administering inhibitory
CC agents to the cell. These telomerase inhibitors can be identified using

CC hstau protein. Antibodies generated against hstau can be used in various
CC immunoassays.

CC Immunoassays.

SQ Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;

Query Match	94.6%;	Score 2971.4;	DB 20;	Length 3190;
Best Local Similarity	97.4%;	Pred. No. 0;		
Matches 3071; Conservative	0;	Mismatches	6;	Indels 76; Gaps 2;

QY	56	cgcgtctctctcgcgtcccgcttcctcttgaaaccgcctcccccgcggcccgcgccgc	115
Db	38	cgcgctccctctcgctcccgcttcctcttgaaaccgcctcccccgcggcccgcgccgc	97
QY	116	gcctctcccaagcgcaactccgcctcttcctcccttcgttcctcttccttcctttt	175
Db	98	gcctctcccaagcgcaactccgcctcttcctcccttcgttcctcttccttcctttt	157
QY	176	tcctctcttccttcctcccttcctcgcgcgcgcacgcgcgcagacgcgcgcgggagcagtc	235
Db	158	tcctctcttccttcctcccttcctcgcgcgcgcacgcgcgcagacgcgcgcgggagcagtc	217
QY	236	ggagcagcagcc-----	247
Db	218	ggagcagcagcagcaggttattaacaccttaacctctcagaactcgaacaagaacacatt	277
QY	248	-----agaaagcataaccctctgtagaactaatgc	280
Db	278	gttcctggaacgcctcttlttaaaaaaagaagacataaccctactgtagaactaatgc	337
QY	281	actgtgcattgaaacttggaaaaaaaccaatgtalaaagcctgttgaaccttactcgcgat	340
Db	338	actgtgcattgaaacttggaaaaaaaccaatgtalaaagcctgttgaaccttactcgcgat	397
QY	341	gcagttccacctataactacaacatgagagggaggtgcttatcccccgaagtactttacc	400
Db	398	gcggtccacctataactacaacatgagagggaggtgcttatcccccgaagtactttacc	457
QY	401	atttcagttccaccttactttatcaagtggaaacttctgttggaggagcagaactttaa	460
Db	458	atttcagttccaccttactttatcaagtggaaacttctgttggaggagcagaactttaa	517
QY	461	tggcgaaggaagaacaagacagcctgcgaaacaacagatgctgtgccaaagcgttgagat	520
Db	518	tggcgaaggaagaacaagacagcctgcgaaacaacagatgctgtgccaaagcgttgagat	577
QY	521	cctgcagaatgagccctctgcagagagagctgtgaggtgaaatggaagagaatccgaagaaga	580
Db	578	cctgcagaatgagccctctgcagagagagctgtgaggtgaaatggaagagaatccgaagaaga	637
QY	581	aaatctcaataaacttgaataaagtcaagtglttgagatlgcacttaaacggaacttgc	640
Db	638	aaatctcaataaacttgaataaagtcaagtglttgagatlgcacttaaacggaacttgc	697
QY	641	tgtagaattcgaagtggtgcccgggagagtggtccacccacatgaaagaacttgtgaccaa	700
Db	698	tgtagaattcgaagtggtgcccgggagagtggtccacccacatgaaagaacttgtgaccaa	757
QY	701	ggttcggttggtggagtttgttggtgggaaggtgaagggaagaagaagatttcaaaaga	760
Db	758	ggttcggttggtggagtttgttggtgggaaggtgaagggaagaagaagatttcaaaaga	817
QY	761	aatgcccgcataagctgttcttgagagagctgaagaagttaacgccccctgtcctgcagttga	820
Db	818	aatgcccgcataagctgttcttgagagagctgaagaagttaacgccccctgtcctgcagttga	877
QY	821	acgagtaaaagccttagaatcaaaaaaagaaaaaaacccatagttaagccaagacaagccc	880
Db	878	acgagtaaaagccttagaatcaaaaaaagaaaaaaacccatagttaagccaagacaagccc	937
QY	881	agaatatgcccaggggatacaatccgattagccgactggtcccagatccagcagcgaaaaaa	940
Db	938	agaatatgcccaggggatacaatccgattagccgactggtcccagatccagcagcgaaaaaa	997

QY	941	ggagaaagagccagagtlacacgctctccacagagcgagggctcccgcgccgcaaggaagt	1000
Db	998	ggaagaagagccagagtlacacgctctccacagagcgagggctcccgcgccgcaaggaagt	1057
QY	1001	tgtgatgcaggttgaaggttggaaaccacactgcagaaggaacggygcaccaacaagaaggt	1060
Db	1058	tgtgatgcaggttgaaggttggaaaccacactgcagaaggaacggygcaccaacaagaaggt	1117
QY	1061	ggccaagcgcaatgcagccggaacaatgcttgagatccttggtttcaaagttcccgcaagc	1120
Db	1118	ggccaagcgcaatgcagccggaacaatgcttgagatccttggtttcaaagttcccgcaagc	1177
QY	1121	gcagcccaaccaaacccgcactcaagtcagaggaagacacccataaagaacaacagggga	1180
Db	1178	gcagcccaaccaaacccgcactcaagtcagaggaagacacccataaagaacaacagggga	1237
QY	1181	tggaaagaaaagtaacctttttgaaacctggtctgtgggagtgaataatlygactagtaataa	1240
Db	1238	tggaaagaaaagtaacctttttgaaacctggtctgtgggagtgaataatlygactagtaataa	1297
QY	1241	agagagtagtcaagatgccttatctaagtcatacagcagctgcctgtggaattctcc	1300
Db	1298	agagagtagtcaagatgccttatctaagtcatacagcagctgcctgtggaattctcc	1357
QY	1301	catgtgtcccggaagtcgcgccaggtctgtagaggttagtlcaaggaacatcacaccaagaatt	1360
Db	1358	catgtgtcccggaagtcgcgccaggtctgtagaggttagtlcaaggaacatcacaccaagaatt	1417
QY	1361	taaccagggcagctccgaatcctgccaagggccaagyltaactgccaatgatagcccggaagt	1420
Db	1418	taaccagggcagctccgaatcctgccaagggccaagyltaactgccaatgatagcccggaagt	1477
QY	1421	gttgatagggggcacctcgcgccacagccgagaccattttaagaataacatctctcagg	1480
Db	1478	gttgatagggggcacctcgcgccacagccgagaccattttaagaataacatctctcagg	1537
QY	1481	ccaagtaccccatgagacctctcacgagaccctctgagcaactgacatacttccagagt	1540
Db	1538	ccaagtaccccatgagacctctcacgagaccctctgagcaactgacatacttccagagt	1597
QY	1541	ccaaggtattccaggttgaatacaaaagacttccccaaaaacaacaagaacgaatttgtatc	1600
Db	1598	ccaaggtattccaggttgaatacaaaagacttccccaaaaacaacaagaacgaatttgtatc	1657
QY	1601	tctatcaattgtctcctctcagccacctctgtatcagccaatgatatcggcaagatgttga	1660
Db	1658	tctatcaattgtctcctctcagccacctctgtatcagccaatgatatcggcaagatgttga	1717
QY	1661	gtccgtgcatgatatagtcgtcgctgaacatcttaaaagtgtgttctgagtttgaccaaca	1720
Db	1718	gtccgtgcatgatatagtcgtcgctgaacatcttaaaagtgtgttctgagtttgaccaaca	1777
QY	1721	aagtaacagagatgcccaagaacaggaacgcgaccaatgtctgtgttgggaggtgtgtgaac	1780
Db	1778	aagtaacagagatgcccaagaacaggaacgcgaccaatgtctgtgttgggaggtgtgtgaac	1837
QY	1781	ctttcttgccatgtaaaccattataaaatcccaacatatatactgaaaataactgaaactgc	1840
Db	1838	ctttcttgccatgtaaaccattataaaatcccaacatatatactgaaaataactgaaactgc	1897
QY	1841	tttgaaaaatttgaatttctgatalactccagltggyccgagagacaacgltggttaagat	1900
Db	1898	tttgaaaaatttgaatttctgatalactccagltggyccgagagacaacgltggttaagat	1957
QY	1901	gttggcagcgacgaggaagacaacagaacaacaaaggaagcggtgttggcc-ggtctgact	1959
Db	1958	gttggcagcgacgaggaagacaacagaacaacaaaggaagcggtgttggccggtctgact	2017
QY	1960	gttgcgtgggttgttgtgtgagtgccaactcgtgtgaaccttggcggtccctaagcaatagcagct	2019
Db	2018	gttgcgtgggttgttgtgtgagtgccaactcgtgtgaaccttggcggtccctaagcaatagcagct	2077

QY 2020 gcctgtgggaagaaggctgcccagcagctggtctcccgggacacccagcatccac 2079
|||||
Db 2078 gcctgtgggaagaaggctgcccagcagctggtctcccgggacacccagcatccac 2137
QY 2080 accctgggcaacctccgtgtttgtgtctttttccctgtgtgaagaagaacgagcag 2139
|||||
Db 2138 accctgggcaacctccgtgtttgtgtctttttccctgtgtgaagaagaacgagcag 2197
QY 2140 acccctctcaagctggtcactcagacacattgagacaacccctgacagccatgcca 2199
|||||
Db 2198 acccctctcaagctggtcactcagacacattgagacaacccctgacagccatgcca 2257
QY 2200 agagagccttgaccggcccccagagctaaaggcaccagagaaaatcaaatgcttcctac 2259
|||||
Db 2258 agagagccttgaccggcccccagagctaaaggcaccagagaaaatcaaatgcttcctac 2317
QY 2260 tcagcgtgacccaacttttctagtgtgccaacggcccccaccacctcctgtcagtaaccacac 2319
|||||
Db 2318 tcagcgtgacccaacttttctagtgtgccaacggcccccaccacctcctgtcagtaaccacac 2377
QY 2320 catcaccactgtctctctcccaacagtgatctgtatctttagtttcaattttctt 2379
|||||
Db 2378 catcaccactgtctctctcccaacagtgatctgtatctttagtttcaattttctt 2437
QY 2380 gattgatatgacacataataaaatttcaattgagaatttctcaattgtatctagtttaa 2439
|||||
Db 2438 gattgatatgacacataataaaatttcaattgagaatttctcaattgtatctagtttaa 2497
QY 2440 tagcacagtttggaactgtctgagacgtacttatcataatctaacccgacaaagatc 2499
|||||
Db 2498 tagcacagtttggaactgtctgagacgtacttatcataatctaacccgacaaagatc 2557
QY 2500 atatcatgtgtatgtgttagacattttatttcaattgactaaaccagagcagtttcag 2559
|||||
Db 2558 atatcatgtgtatgtgttagacattttatttcaattgactaaaccagagcagtttcag 2617
QY 2560 tgatgcaaatgtgtgccctcgtgtcagcgtgaaacagtcctggaacttcaaaaacctg 2619
|||||
Db 2618 tgatgcaaatgtgtgccctcgtgtcagcgtgaaacagtcctggaacttcaaaaacctg 2677
QY 2620 aataagctcccaagatgtataaatgagcaatttaggaattttaacttagatgac 2679
|||||
Db 2678 aataagctcccaagatgtataaatgagcaatttaggaattttaacttagatgac 2737
QY 2680 attggttccattttttatttattttttgttaatgcaaacagagacttaaatgaac 2739
|||||
Db 2738 attggttccattttttatttattttttgttaatgcaaacagagacttaaatgaac 2797
QY 2740 ttgatctcgttttaaaagattataaaaaacattgtgtatctatataatgctcttga 2799
|||||
Db 2798 ttgatctcgttttaaaagattataaaaaacattgtgtatctatataatgctcttga 2857
QY 2800 ggaactagcttccactacactacagatagatcctcatgtagtcacataataaacctgcag 2859
|||||
Db 2858 ggaactagcttccactacactacagatagatcctcatgtagtcacataataaacctgcag 2917
QY 2860 agtgatttccagagtgctcgatactgttaattacatctccattagggctgaaaagatg 2919
|||||
Db 2918 agtgatttccagagtgctcgatactgttaattacatctccattagggctgaaaagatg 2977
QY 2920 accctagttctgtatatacagctgtgttgccttttgatgtgtgttactgtacacagaagt 2979
|||||
Db 2978 accctagttctgtatatacagctgtgttgccttttgatgtgtgttactgtacacagaagt 3037
QY 2980 tgtgcaactgagcctcgtcgtgtgtccgfatggaanaacctgtagccctgcgagttaat 3039
|||||
Db 3038 tgtgcaactgagcctcgtcgtgtgtccgfatggaanaacctgtagccctgcgagttaat 3097
QY 3040 actgcttccattcaattgtttaacgctggaattttctcccatggaatgtaagtaaaact 3099
|||||
Db 3098 actgcttccattcaattgtttaacgctggaattttctcccatggaatgtaagtaaaact 3157
QY 3100 aagtgtttgtcatcaataaatggttaatactaaa 3132

Db 3158 aagtgtttgtcatcaataaatggttaatactaaa 3190
|||||
RESULT 4
293322
ID Z93322 standard; cDNA; 3506 BP.
XX
AC Z93322;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 409..2142
FT /*tag= a
FT /product= Stauflen
CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI Mouland AJ;
XX
XX WPI: 2000-246924/22.
DR P-PSDB; Y83023.
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96bp; English.
XX
CC Stauflen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC stauflen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian stauflen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauflen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;
Query Match 92.2%; Score 2896; DB 21; Length 3506;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 cagaaagcataaccctactgtagaactaaatgcaactgtgcatgaaacttgaaaaaac 306
|||||
Db 611 cagaaagcataaccctactgtagaactaaatgcaactgtgcatgaaacttgaaaaaac 670

QY	307	caatgataaagcctgttgacccttactctcggatgacgtccacctataactacaacatga	366
Db	671	caatgtataaagcctgttgacccttactctcggatgacgtccacctataactacaacatga	730
QY	367	gaggaagtgcttatcccccgaagtgacttttaaccatttccagttccaccttactatc	426
Db	731	gaggaagtgcttatcccccgaagtgacttttaaccatttccagttccaccttactatc	790
QY	427	aagtggaaacttctgttgggggacagcaatttaatgccaaggaaagacaagacagctg	486
Db	791	aagtggaaacttctgttgggggacagcaatttaatgccaaggaaagacaagacagctg	850
QY	487	cgaaacacgatgtctgtcccaaaagcgttgagatccttcgagaatgagccctgcagaga	546
Db	851	cgaaacacgatgtctgtcccaaaagcgttgagatccttcgagaatgagccctgcagaga	910
QY	547	ggctggaggtgaatggaagaaatccgaagaagaaatcctaataatctgaataagtc	606
Db	911	ggctggaggtgaatggaagaaatccgaagaagaaatcctaataatctgaataagtc	970
QY	607	aagtgtttgagattgcaacttaaacggaacttgctgtgaatttcgaagtggtgccggaga	666
Db	971	aagtgtttgagattgcaacttaaacggaacttgctgtgaatttcgaagtggtgccggaga	1030
QY	667	gtggcccaccccacatgaaagaacttgtgaccaagtttcggttggggagttgttggggg	726
Db	1031	gtggcccaccccacatgaaagaacttgtgaccaagtttcggttggggagttgttggggg	1090
QY	727	aagtgtaagggaaaaagcaagaagaatttcaagaagaaatgcccgcataagctgttcttgag	786
Db	1091	aagtgtaagggaaaaagcaagaagaatttcaagaagaaatgcccgcataagctgttcttgag	1150
QY	787	agctggaagaagttacccgcccctgctgcagtltgaacgagtaaaagcctagaatcaaaaaa	846
Db	1151	agctggaagaagttacccgcccctgctgcagtltgaacgagtaaaagcctagaatcaaaaaa	1210
QY	847	aaacaaaaccatagttcaagccaacagacagcccgaaatgycagggatcaatccga	906
Db	1211	aaacaaaaccatagttcaagccaacagacagcccgaaatgycagggatcaatccga	1270
QY	907	ttagccgactgcccagatccagcagggcaaaaaaggagaaggagccagatcacgcctcc	966
Db	1271	ttagccgactgcccagatccagcagggcaaaaaaggagaaggagccagatcacgcctcc	1330
QY	967	tcacagagcgaagcctcccgcgccgcaaggagtttgtatgcaagtgaagtttggaaacc	1026
Db	1331	tcacagagcgaagcctcccgcgccgcaaggagtttgtatgcaagtgaagtttggaaacc	1390
QY	1027	acaactgcagaaggaacccggaaccaacaagaagttgcccgaagcgcgaatgcaagccga	1086
Db	1391	acaactgcagaaggaacccggaaccaacaagaagttgcccgaagcgcgaatgcaagccga	1450
QY	1087	tgctgagatccttgtttcaaaagttcccgcaagcgcgcccaccaaacccgcactcaagt	1146
Db	1451	tgctgagatccttgtttcaaaagttcccgcaagcgcgcccaccaaacccgcactcaagt	1510
QY	1147	cagagggagaagacaccccataaagaacaacgaagggaatggaagaaagtaacctttttgaac	1206
Db	1511	cagagggagaagacaccccataaagaacaacgaagggaatggaagaaagtaacctttttgaac	1570
QY	1207	ctggtcttgggatgaaaaatgggactagtataaagagatgatttcagagatgcctatc	1266
Db	1571	ctggtcttgggatgaaaaatgggactagtataaagagatgatttcagagatgcctatc	1630
QY	1267	taagtcacagcagctgcctgtgtgaatttcttccatgtgtcccgaagtgcgccagctg	1326
Db	1631	taagtcacagcagctgcctgtgtgaatttcttccatgtgtcccgaagtgcgccagctg	1690
QY	1327	taggaagttagtcaaggaacatcacaccaaagattttaaccaagggcagctccgaatcctgcca	1386
Db	1691	taggaagttagtcaaggaacatcacaccaaagattttaaccaagggcagctccgaatcctgcca	1750

QY	1387	aggccacggtaaactgccaatgataagcccgagaggtgtgtataggggcaacctgcgccacag	1446
Db	1751	agggcacgtaactgccaatgataagcccgagaggtgtgtataggggcaacctgcgccacag	1810
QY	1447	ccgagacacattttaagaataacatctcttcagggccaagtaccccatggaacctctcaca	1506
Db	1811	ccgagacacattttaagaataacatctcttcagggccaagtaccccatggaacctctcaca	1870
QY	1507	gacctctgagcaactggaatactcttccagagtcacaggaatccaggttgaatacaaaag	1566
Db	1871	gacctctgagcaactggaatactcttccagagtcacaggaatccaggttgaatacaaaag	1930
QY	1567	acttccccaaaaaacaacaagaacgaatttgtatctcttatacaattgctccctcagccac	1626
Db	1931	acttccccaaaaaacaacaagaacgaatttgtatctcttatacaattgctccctcagccac	1990
QY	1627	ctctgatcagccaatggtatcgggcaagaagtgtgagtlccctgcataatgagctgcgtga	1686
Db	1991	ctctgatcagccaatggtatcgggcaagaagtgtgagtlccctgcataatgagctgcgtga	2050
QY	1687	acatcttaaaagtgtgtctgtgaagttgagccaacaaagtaacagatgccaagaacagaa	1746
Db	2051	acatcttaaaagtgtgtctgtgaagttgagccaacaaagtaacagatgccaagaacagaa	2110
QY	1747	acggaccaaagtctgtgtgtgaggagtgctgtaaccttctctggccaatgaacctataaa	1806
Db	2111	acggaccaaagtctgtgtgtgaggagtgctgtaaccttctctggccaatgaacctataaa	2170
QY	1807	atcccaacataataactgaataactgcttgaanaatttgaattctgatacc	1866
Db	2171	atcccaacataataactgaataactgcttgaanaatttgaattctgatacc	2230
QY	1867	tcacagtgggcccagagagacacggtggtaaaagtgtggcagcagcagggaagaacaacag	1926
Db	2231	tcacagtgggcccagagagacacggtggtaaaagtgtggcagcagcagggaagaacaacag	2290
QY	1927	aaacacaaggaggcggctgtgcccggctggaactgtctggggttgtgtgatggccact	1986
Db	2291	aaacacaaggaggcggctgtgcccggctggaactgtctggggttgtgtgatggccact	2350
QY	1987	cggtagcctggcggtccctaagcaatagcagctgctgtgggaagaaggcgtgccagc	2046
Db	2351	cggtagcctggcggtccctaagcaatagcagctgctgtgggaagaaggcgtgccagc	2410
QY	2047	cagctggtctcccgggacacccagcagatccacaccccttggcaccctccggttggctct	2106
Db	2411	cagctggtctcccgggacacccagcagatccacaccccttggcaccctccggttggctct	2470
QY	2107	tttttccctgtgtgaagaagaacggcagcagcccttctcaagctggtccactcaga	2166
Db	2471	tttttccctgtgtgaagaagaacggcagcagcccttctcaagctggtccactcaga	2530
QY	2167	cacatgaggacaacaccctggacagccatgcccagagagagccttgaacggccccagagc	2226
Db	2531	cacatgaggacaacaccctggacagccatgcccagagagagccttgaacggccccagagc	2590
QY	2227	taaaagcaccagagaaaatcaaatgcttcctactcagcgtgacccaacttctctagtg	2286
Db	2591	taaaagcaccagagaaaatcaaatgcttcctactcagcgtgacccaacttctctagtg	2650
QY	2287	ccacggccccacacacctctcgtcagtaaccaacacatcaccaactgcttctctccaacag	2346
Db	2651	ccacggccccacacacctctcgtcagtaaccaacacatcaccaactgcttctctccaacag	2710
QY	2347	tgatctgatctctagtttcattatcttctcttgatgatatgacactataaaattt	2406
Db	2711	tgatctgatctctagtttcattatcttctcttgatgatatgacactataaaattt	2770
QY	2407	catttgagaatttctcaattgtatatctagttaaatagcacagtttggaactgtctgaga	2466
Db	2771	catttgagaatttctcaattgtatatctagttaaatagcacagtttggaactgtctgaga	2830
QY	2467	ctgaacttatacaataactcaaccgcaaaagatcataatcaatgtgtatgtgttagacatt	2526

|||||
Db 2831 ctgactttatcataatctaaccgacaagatcatatccatgtgtatgtgtagacatt 2890
QY 2527 ttatttcattgactaaccagcaggttcagtgatgcaaatgtgtgccctctgttc 2586
Db 2891 ttatttcattgactaaccagcaggttcagtgatgcaaatgtgtgccctctgttc 2950
QY 2587 agctgaacagtcctgtgactttcaaaaaaccttgaataagtcctccacagttgtataatc 2646
Db 2951 agctgaacagtcctgtgactttcaaaaaaccttgaataagtcctccacagttgtataatc 3010
QY 2647 ggacaatttagaattttaaacctttagatgatacatttggtccatttttattcatttt 2706
Db 3011 ggacaatttagaattttaaacctttagatgatacatttggtccatttttattcatttt 3070
QY 2707 attttgttaattgaaacagagacttaaatgaactttgatctctgttttaagaattataa 2766
Db 3071 attttgttaattgaaacagagacttaaatgaactttgatctctgttttaagaattataa 3130
QY 2767 aaaacattgtgtatctatcatatagtcctcttgaggacttagcttccactactacagga 2826
Db 3131 aaaacattgtgtatctatcatatagtcctcttgaggacttagcttccactactacagga 3190
QY 2827 tatgatctccatgtagtcacataaaacctgcagagtgattttccagagtgctcgatactg 2886
Db 3191 tatgatctccatgtagtcacataaaacctgcagagtgattttccagagtgctcgatactg 3250
QY 2887 ttaatacatctccattagggctggaagaatgaacctacgtttctgtatacagctgtgt 2946
Db 3251 ttaatacatctccattagggctggaagaatgaacctacgtttctgtatacagctgtgt 3310
QY 2947 gctttgatgtgtgttactgtacacagaagtgtgtgacatgaggtcgtggtgtcc 3006
Db 3311 gctttgatgtgtgttactgtacacagaagtgtgtgacatgaggtcgtggtgtcc 3370
QY 3007 gtatggaacacctgtgtagccctgcgagttaagtactgtctccattcattgtttaacgtg 3066
Db 3371 gtatggaacacctgtgtagccctgcgagttaagtactgtctccattcattgtttaacgtg 3430
QY 3067 aattttctcccatggaatgtgaagtaaaacttaagtgtttgtcatcataataatgttaat 3126
Db 3431 aattttctcccatggaatgtgaagtaaaacttaagtgtttgtcatcataataatgttaat 3490
QY 3127 actaaaaaaaaa 3142
Db 3491 actaaaaaaaaa 3506
RESULT 5
293327
ID 293327 standard; cDNA; 3260 BP.
XX
AC 293327;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human staufen cDNA.
XX
KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; ; human; ss.
OS
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 409..1311
FT /*tag=a
FT /product= Staufen protein
XX CA2238656-A1.
XX
PD 22-NOV-1999.
XX

PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y63108.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1b; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
SQ Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;
Query Match 92.0%; Score 2890.2; DB 21; Length 3260;
Best Local Similarity 94.6%; Pred. No. 0;
Matches 3086; Conservative 0; Mismatches 53; Indels 124; Gaps 4;
QY 1 acttcctgcgggctgcgggcgccttgagcgtcttcacgcttttgccggcgctgcgct 60
Db 1 acttcctgcgggctgcgggcgccttgagcgtcttcacgcttttgccggcgctgcgct 60
QY 61 ctctctcgctccgcgttcctttgacgcgcctcccccccgccggcgccgcgcctc 120
Db 61 ctctctcgctccgcgttcctttgacgcgcctcccccccgccggcgccgcgcctc 120
QY 121 ctccagcgccactccgcctctccctccctcg----- 153
Db 121 ctccagcgccactccgcctctccctccctcgagcagccagcagcgagtgag 180
QY 154 -----tccttctctctc 168
Db 181 ttgctctgtcggccaggtgagtgagtgagtgatctcgctcactgcaacctca 240
QY 169 cct-----tttccctctctccctc----- 192
Db 241 cctccaggtcagcgatttcccaacttcagcctcccgataagctgagattaaggagttt 300
QY 193 -----cctgcgcgcaacgcgcccagagcgcgcgcggggagagctcgag 239
Db 301 attaaccaacttaacctctcagaactgaacaagaacaacatgttcttggaagccctct 360
QY 240 cagcagccagaagaagcataccctactgtagaactaaatgacactgtgcatgaaacttga 299
Db 361 ttaaaaaaagaagcataaccctactgtagaactaaatgacactgtgcatgaaacttga 420
QY 300 aaaaaaccaatgtataagcctgttgacccttactctcgatgagtcacactataactac 359
|||||

Db 421 aaaaaccaatgtataagcctgttgacccttactctcgatgagt---cctataactac 477
QY 360 aacatgagagagtgctctatccccgagtlacttltaccatttccagttccacctta 419
Db 478 aacatgagagagtgctctatccccgagtlacttltaccatttccagttccacctta 537
QY 420 cttaataagtggaacttctgtggagagacagcaatttaatgycaaagaaagacaaga 479
Db 538 cttaataagtggaacttctgtggagagacagcaatttaatgycaaagaaagacaaga 597
QY 480 caggtgcgaacaacacgatagtctgtgcgaagcggttgagatctctgcagaatgagccctg 539
Db 598 caggtgcgaacaacacgatagtctgtgcgaagcggttgagatctctgcagaatgagccctg 657
QY 540 ccagaagagctggaagtgatgaaatgcgaagaagaataatctcaataaattgaa 599
Db 658 ccagaagagctggaagtgatgaaatgcgaagaagaataatctcaataaattgaa 717
QY 600 ataagtcgaagtgttgagatgcaacttaaacggaactgcctgtgaaatttcgaggtgccc 659
Db 718 ataagtcgaagtgttgagatgcaacttaaacggaactgcctgtgaaatttcgaggtgccc 777
QY 660 cgggagagtgcccaaccacatgaagaacttltgaccaaagtttcggttgaggagttt 719
Db 778 cgggagagtgcccaaccacatgaagaacttltgaccaaagtttcggttgaggagttt 837
QY 720 gtggggaaggttgaaaggaaagcaagaagatttcaagaanaaatgcccagatagctgtt 779
Db 838 gtggggaaggttgaaaggaaagcaagaagatttcaagaanaaatgcccagatagctgtt 897
QY 780 ctgagagagctggaagaagttaaccgccccctgctgcagtltgaacgagttaaagcctagaatc 839
Db 898 ctgagagagctggaagaagttaaccgccccctgctgcagtltgaacgagttaaagcctagaatc 957
QY 840 aaaaagaaaaaaccccatagtcgaagccacagacaagaagcccaagaatatgcccaggygatac 899
Db 958 aaaaagaaaaaaccccatagtcgaagccacagacaagaagcccaagaatatgcccaggygatac 1017
QY 900 aatccgattagccgaactgyccccagatccagcagycaaanaaggaagagccagagttac 959
Db 1018 aatccgattagccgaactgyccccagatccagcagycaaanaaggaagagccagagttac 1077
QY 960 acgctcctcacagagcgagcgctcccgcgccgagggagttctgtatgcaagtgaagtt 1019
Db 1078 acgctcctcacagagcgagcgctcccgcgccgagggagttctgtatgcaagtgaagtt 1137
QY 1020 ggaaccacacatgcaagaaggaaagggcaccaacaagaagtgycacaagcgaatgcaagcc 1079
Db 1138 ggaaccacacatgcaagaaggaaagggcaccaacaagaagtgycacaagcgaatgcaagcc 1197
QY 1080 gagaacatgctgtagatccttgttccaagttcccgagcgagcccaaccaaacccgca 1139
Db 1198 gagaacatgctgtagatccttgttccaagttcccgagcgagcccaaccaaacccgca 1257
QY 1140 ctcaagtcagaggaagaagacaccccaataaagaaacagggagatgaaagaaagtaacctt 1199
Db 1258 ctcaagtcagaggaagaagacaccccaataaagaaacagggagatgaaagaaagtaacctt 1317
QY 1200 tttaaacctgctctggygagtaagaatggyagtagtaataaagagagtagttcagatg 1259
Db 1318 tttaaacctgctctggygagtaagaatggyagtagtaataaagagagtagttcagatg 1377
QY 1260 cctatactaagtcacagcagctgcctgtgtgaattcttcccatgtgtgccgaggtgcgc 1319
Db 1378 cctatactaagtcacagcagctgcctgtgtgaattcttcccatgtgtgccgaggtgcgc 1437
QY 1320 caggtctgtgaggttagtcaagagacatcacaccaagaattttaaccagggcagctccgaat 1379
Db 1438 caggtctgtgaggttagtcaagagacatcacaccaagaattttaaccagggcagctccgaat 1497
QY 1380 cctgccaaagccaagytacactgccaatagcccgagagttgttatatggyggcacctcg 1439
Db 1498 cctgccaaagccaagytacactgccaatagcccgagagttgttatatggyggcacctcg 1557

QY 1440 cccacagccgagaccattttaagaataacatctcttcagggccaagcgtlaccocatgagact 1499
Db 1558 cccacagccgagaccattttaagaataacatctcttcagggccaagcgtlaccocatgagact 1617
QY 1500 ctcaagagaccctctgagcaactgtagtatacttccagagttccaggagttccaggttgaa 1559
Db 1618 ctcaagagaccctctgagcaactgtagtatacttccagagttccaggagttccaggttgaa 1677
QY 1560 tacaagaacttccccaaaaacaacaagaacgaaatttgtatctcttaatcaattgtcctct 1619
Db 1678 tacaagaacttccccaaaaacaacaagaacgaaatttgtatctcttaatcaattgtcctct 1737
QY 1620 cagccacctctgatacagcatggtatccggcaaggatgtgagtcctgcatagtatgct 1679
Db 1738 cagccacctctgatacagcatggtatccggcaaggatgtgagtcctgcatagtatgct 1797
QY 1680 gcgctgaacatcttaaaagtgtgtctcctgagttgaccaaacaagtlacagagatgccaaga 1739
Db 1798 gcgctgaacatcttaaaagtgtgtctcctgagttgaccaaacaagtlacagagatgccaaga 1857
QY 1740 acaggaacaggaaccaatgtctgtgtgtggaggtgtgtgaacctttctgccaatgacca 1799
Db 1858 acaggaacaggaaccaatgtctgtgtgtggaggtgtgtgaacctttctgccaatgacca 1917
QY 1800 ttataaaatcccaacatatatacttgaanaatactgaaactgtcttgaanaatttgaaatttc 1859
Db 1918 ttataaaatcccaacatatatacttgaanaatactgaaactgtcttgaanaatttgaaatttc 1977
QY 1860 tgatacctccagttggccgagagacaacggtggtlaaagatgtggcagcagcagggaaag 1919
Db 1978 tgatacctccagttggccgagagacaacggtggtlaaagatgtggcagcagcagggaaag 2037
QY 1920 acaacagaacaacaaggagggcggtgtggccggtgtagctgtgtgggttgtgtgat 1979
Db 2038 acaacagaacaacaaggagggcggtgtggccggtgtagctgtgtgggttgtgtgat 2097
QY 1980 ggcacctcgtgtgacctgtgcygttccctacgcaatagcagctgcctgtggggaagaaggct 2039
Db 2098 ggcacctcgtgtgacctgtgcygttccctacgcaatagcagctgcctgtggggaagaaggct 2157
QY 2040 gccacagcagctgtgttctcccgggagacacagcagatccacacccctgggcacctccgtgtt 2099
Db 2158 gccacagcagctgtgttctcccgggagacacagcagatccacacccctgggcacctccgtgtt 2217
QY 2100 tgytcttlttccctgtgtgtaagaagaagaacgycacagaccccttccaagctgtgtc 2159
Db 2218 tgytcttlttccctgtgtgtaagaagaagaacgycacagaccccttccaagctgtgtc 2277
QY 2160 actcagacacatggygacaacacctgagacagccatgycagagagaggccttgaaccgccc 2219
Db 2278 actcagacacatggygacaacacctgagacagccatgycagagagaggccttgaaccgccc 2337
QY 2220 ccagagctaaagcaccagagaanaatccaatgtcttccctactcagcgctgacccaactttc 2279
Db 2338 ccagagctaaagcaccagagaanaatccaatgtcttccctactcagcgctgacccaactttc 2397
QY 2280 tagtgtgccaagggcccaacacactcctgtagtacccaacacatcaaccaactgttctct 2339
Db 2398 tagtgtgccaagggcccaacacactcctgtagtacccaacacatcaaccaactgttctct 2457
QY 2340 ccaacagtlatctgtatctttagttcatatttcttlttgatgtatatgacactata 2399
Db 2458 ccaacagtlatctgtatctttagttcatatttcttlttgatgtatatgacactata 2517
QY 2400 aaatttcatltgagaatttctcaattgtatctagttaaatagcacagtttggaacttg 2459
Db 2518 aaatttcatltgagaatttctcaattgtatctagttaaatagcacagtttggaacttg 2577
QY 2460 tctgagactgacttatacataatcttaaccgacaagaatcataatcatgtgtatgtgtt 2519
Db 2578 tctgagactgacttatacataatcttaaccgacaagaatcataatcatgtgtatgtgtt 2637

QY	2520	agacatttatttcatttgactaacccagagacgttcacgtgatgcaaatgtgtgccct	2579
Db	2638	agacatttatttcatttgactaacccagagacgttcacgtgatgcaaatgtgtgccct	2697
QY	2580	ctgtgtcagctggaacacgtccctgagacttcaaaaaccttgataagtctccacagttgt	2639
Db	2698	ctgtgtcagctggaacacgtccctgagacttcaaaaaccttgataagtctccacagttgt	2757
QY	2640	ataaattggacaatttaggaattttaacctttagatgatcatttggttccattttattt	2699
Db	2758	ataaattggacaatttaggaattttaacctttagatgatcatttggttccattttattt	2817
QY	2700	catttttattttgtgttaatgcaaacaggaacttaaatgaactttgatctctgttttaaga	2759
Db	2818	catttttattttgtgttaatgcaaacaggaacttaaatgaactttgatctctgttttaaga	2877
QY	2760	ttatlaaaaaacatgtgtatctatacatatgctcttggagacttagctttcacataac	2819
Db	2878	ttatlaaaaaacatgtgtatctatacatatgctcttggagacttagctttcacataac	2937
QY	2820	tacagatatgatctcccatgttagtccatataaaccttgcagagtattttccagagtctc	2879
Db	2938	tacagatatgatctcccatgttagtccatataaaccttgcagagtattttccagagtctc	2997
QY	2880	gatactgttaattacatctccatttagggctgaaagaatgacctagcttctgtatacag	2939
Db	2998	gatactgttaattacatctccatttagggctgaaagaatgacctagcttctgtatacag	3057
QY	2940	ctgtgtgtcctttgtagtgtgtgttactgttacacagaagtgtgtgcaactgaggtctgcgt	2999
Db	3058	ctgtgtgtcctttgtagtgtgtgttactgttacacagaagtgtgtgcaactgaggtctgcgt	3117
QY	3000	gtgtgtccgtatggaanaacctgtgtagccctgcgagtttaagtaactgtcttccattcatgttt	3059
Db	3118	gtgtgtccgtatggaanaacctgtgtagccctgcgagtttaagtaactgtcttccattcatgttt	3177
QY	3060	acgctggaatttttctcccatgtgaatgttaagtaaaacttaagtgtttgtcatcataaa	3119
Db	3178	acgctggaatttttctcccatgtgaatgttaagtaaaacttaagtgtttgtcatcataaa	3237
QY	3120	tgtgtaatactaaaaaataaa 3142	
Db	3238	tgtgtaatactaaaaaataaa 3260	

RESULT	6	
ID	Z93328	standard; cDNA; 2859 BP.
XX		
AC	Z93328;	
XX		
DT	04-JUL-2000	(first entry)
XX		
DE	Mouse staufen cDNA.	
XX		
KW	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
XX	acquired immune deficiency syndrome; protease; mouse; ss.	
XX		
OS	Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	325..1788
FT		/*tag= a
FT		/product= Staufen protein
XX		
PN	CA2238656-A1.	
XX		
PD	22-NOV-1999.	
XX		
PF	22-MAY-1998;	98CA-2238656.
XX		
22-MAY-1998;	98CA-2238656.	

XX (UYMO-) UNIV MONTREAL.
PA
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;
PI Moulard AJ;
XX
DR WPI: 2000-246924/22.
DR P-PSDB; Y83024.
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS
XX
XX Claim 4; Figure 1c; 96pp; English.
CC
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
XX
SQ Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;

Query Match	43.2%	Score 1356.2;	DB 21;	Length 2859;
Best Local Similarity	84.7%;	Pred. No. 0;		
Matches 1598; Conservative	0;	Mismatches 268;	Indels 21;	Gaps 6;

QY	1	acttctcgcgagctgcggcgctgagcgctcttcaagcgtttgcgcg--ggtctgcg	58
	24	acttctcgcagcgtgcggcccgagcgctcttcaagcgtttgcgtgcgtgcgc	83
QY	59	gtctctcgcgctccgcgttctcttgaacgcctccccccgcgcgcgcgcgcgc	118
Db	84	gtctgtgtgcgtccccccttctctgag-----ccccgcgtgcgcgcgcgcgc	133
QY	119	tctctcaagcgccactccgcctcttccctccctctgcctctctctcccttttcc	178
Db	134	ttcgctccgcgcactccgcctcttccctctctgtgcgcgccttttctctgcgcgtct	193
QY	179	ttcttcttccctccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	237
Db	194	cacttgcttcttcaacctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	253
QY	238	agcagcagccagaagcataaccctactctgtagaactaaatgcactgtgcataacttg	297
Db	254	ggaagcagccagaagatagcttcttcaacatltgagctcaatgcactgtgtgaaacttg	313
QY	298	gaaaaaaacccaatgtataagcctgtttgacccttactctcggatgcagtcaccctataact	357
Db	314	aaagaaaacccaatgtataagcccgltgaaacctcaactctcgatgcagtcaccctataagct	373
QY	358	acaacatgagagaggtgtctatcccccgaggtactttaaccatltccagltccacct	417
Db	374	atgycatgcgtgaggtgtcctatccccccagatacctttaaccatltccagltccacct	433
QY	418	tactttatacaagtgaaacttctgtgtagagacagcaatttaatgycaaaaggaaagacaa	477
Db	434	tacttacaagtgagctctccgltgtagcagacagcagtttaatgycaaaaggaaagatga	493
QY	478	gacagcgtgcgaaacacgatgtgtctgccaagcgtltgagatcctgcagaatgagcccc	537
Db	494	gaccacccgtlgaacacagatgccctgtgccctgtgcgtgagagactctgcagagatgaacccc	553

QY 538 tgcagagagctgaggtgaatggaagaaatccgaagaagaatctcaataatctg 597
||||| ||| ||||||| ||||||| ||| ||||||| ||||||| |||
Db 554 tgcagaaagttgaggtaaatggaagaaagcagaggaagaaacctcaataatcgg 613
QY 598 aaataagtcaggttctgagattgcacttaacggaacttgcctgtgaatttcgagtg 657
||||| ||||||| ||||||| ||| ||| ||| ||||||| ||||||| |||
Db 614 aaataagccaagttctgaaattgcgtgaaagcggaatttgcctgtgaattttgagtg 673
QY 658 cccgggagagtggcccaaccacatagaaacttctgtgccaagtttcggttgggagt 717
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
Db 674 cccgggagagtggcccaaccacatagaaacttctgtgaccaggtttcagttgggaat 733
QY 718 ttgtggggaaggtgaagggaagcaagaagatttcaagaaaatgcgcgcatagtctg 777
||||| ||||||| ||||||| ||||||| ||| ||| ||| ||| ||| |||
Db 734 ttgtagggaaggaagggaagaaagcaagaagatctccaagaagaatgcggccagggctg 793
QY 778 ttcttgagagcttgaagaagttaccgccccctgctgcagtltgaacgagtaagccctagaa 837
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 794 ttctggagcagcttgaaggagctgccaacccctcctgtgtgtgagcgagtgaaagcccaagaa 853
QY 838 tcaaaaagaaaacaaaccatagtcaagccacagacaagcccaagaatatggccaggga 897
||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 tcaagaaagaaaagtcagcccaacctgca-----agacagcccccgataltatggccaaggga 907
QY 898 tcaatccgatttagccgaactggccagatccagcaggcaaaaaagagagaagagccagagt 957
||||| ||||||| ||| ||| ||| ||||||| ||||||| ||||||| |||
Db 908 tgaatccctattagtagacttgacacagatccagcagycagcaaaaaagagagaagagccagagt 967
QY 958 acaagctcctcacagagcgaagcctcccgcgccgagggagttgtgatgcaggtgaag 1017
||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 968 acatgtcctctacagaaagaggtcttccacgctcgcaaggagttgtgatgcaggtaaag 1027
QY 1018 ttggaaccacacactgcaagaaggaacgagccaccaagaagagtgccaagcgcaatgcag 1077
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1028 ttgggcatacacactgcaagaagagtggtgtaccataagaagtgccaagcgtaatgtctg 1087
QY 1078 ccgagaacatgctgagagatccttgttttcaaaagtcccgagcgccaccaaccaaccg 1137
||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1088 ctgagaacatgctgagagatccttgggttcaaaagttccccaagcgccagctgccaagccag 1147
QY 1138 cactcaagtcagagagaagacaaccataaagaaaccagggagtgaagaaagtaacct 1197
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1148 cactcaaatcagaagaagaagactccagtaaaagaaaccaggaagcgaagaaagtaacct 1207
QY 1198 ttttgaacctggtctggggatgaaaatgggactagtaataaagagagatgattcagaa 1257
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1208 ttttgaacctagccctggggatgaaaatgtaactagtaacaagagcagagagttcagaa 1267
QY 1258 tgccttatctaagtcacagcagctgctgtggaattcttcccatgtgccgagtg 1317
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1268 tgccttatcttagccatcagcagctgccaagcggaattctcccatgtgtgccgaagtg 1327
QY 1318 cccaggtgttagagttagtcagagacatcacaccaaaagattttaccagggcagctccga 1377
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1328 cccaggtgttcgggttagtcagagacacccaacaaagattttaccagggcagctccaa 1387
QY 1378 atcctgccaagggccaagtaactgccaatgataggcccagaggtgtgtatggggcacct 1437
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1388 atcctgccaagggccaagtaactgccaatgataggcccagaggtgtgtatggggcacct 1447
QY 1438 cgcccaagcggaagccattttaaagaatacatctctcagggccagtaaccatgagc 1497
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1448 cgcccaagcggaagccattttaaagaatacatctctcagggccagtaaccatgagc 1507
QY 1498 ctctcagagacctctgagcaactgactatcttccagagtcagggattccaggttg 1557
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1508 ctgcactagaccctctgagcaactgtactaccttccagagcccgaggtatccaggtg 1567
QY 1558 aatacaaaagacttccccaaaaaacacaagaagaaattgtatatcttatacaattgtcct 1617
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1568 aatacaaaagatttccccaaagaacaacaagaagagtggtatatcttcatcaactgtcct 1627

QY 1618 ctcaagccacctgtgatcaagccatggtatcgccaaggatgtggagtcctgcgatgatgg 1677
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1628 caacagcgccctctgcgtcaagtcatgcatcgccaaggatgtggagtcctgcgatgatgg 1687
QY 1678 ctgcgtgaaacatcttaaaagtgtgtctgtgagtggaaccaacaagatacagagatggcaa 1737
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1688 ctgcactgaacattttaaagctgtgtctgtgagtggaaccaacagagcagagatggcaa 1747
QY 1738 gaacaggaacagcgaacatgtctgtgtgtggaggtgctgaacctttctgccaatgaac 1797
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1748 gaacaggaaatggaaccggttcaagcgtgcgggaggtgtgaacctttctgccaacaac 1807
QY 1798 cattataaatcccaacatatatactgaaatactga-aactgcttggaaaaatttggaa 1856
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1808 cattataaaa-cccaacatatatactgaaataactgagaactgtcttggaaaaatttggaa 1866
QY 1857 ttctgatacctccagtgggccgagaga 1883
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1867 atctgataactccagtgggccaagaca 1893

RESULT 7
X39991
ID X39991 standard; DNA; 1127 BP.
XX
AC X39991;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Oyata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PS
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 625-626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 1127 BP, 349 A; 265 C; 307 G; 203 T; 3 other;

Query Match 35.2%; Score 1107.2; DB 20; Length 1127;
Best Local Similarity 99.4%; Pred. No. 1e-270;
Matches 1120; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

OY 445 gagacagcaatttaatgcaagaagacaagacagctgcgaaacacgattgctg 504
Db 1 gagacagcaatttaatgcaagaagacaagacagctgcgaaacacgattgctg 60
OY 505 ccaagcgttgaggtcctctgcagatgagccctgcagagagctgagtgatgaa 564
Db 61 ccaagcgttgaggtcctctgcagatgagccctgcagagagctgagtgatgaa 120
OY 565 gagatccgaaagaagaaatctcaataatctgaataagtcaagtgtttgagattgcac 624
Db 121 gagaatccgaaagaagaaatctcaataatctgaataagtcaagtgtttgagattgcac 180
OY 625 ttaaacggaacttgctgtgaatttcgaaggtgcccgggagagtggcccaaccacatga 684
Db 181 ttaaacggaacttgctgtgaatttcgaaggtgcccgggagagtggcccaaccacatga 240
OY 685 agaacttctgaaccaaggttctcggttgaggagttgtgagggaaggtgaaggaaagca 744
Db 241 agaacttctgaaccaaggttctcggttgaggagttgtgagggaaggtgaaggaaagca 300
OY 745 agaagattcaagaagaaaatgccgcacatagctgttcttgagagctgaagaagttaccgc 804
Db 301 agaagattcaagaagaaaatgccgcacatagctgttcttgagagctgaagaagttaccgc 360
OY 805 ccctgcctgcagttgaacgagtaaagccttagaatcaaaaaagaaacaaaccatagtca 864
Db 361 ccctgcctgcagttgaacgagtaaagccttagaatcaaaaaagaaacaaaccatagtca 420
OY 865 agccacagacaagcccagaaatatgcccaggggatcatcaccgattagccgactggcccaga 924
Db 421 agccacagacaagcccagaaatatgcccaggggatcatcaccgattagccgactggcccaga 480
OY 925 tccagcaggcacaaaaaggaagaagcagcagtlacacgctcctcacagagcgagcctcc 984
Db 481 tccagcaggcacaaaaaggaagaagcagcagtlacacgctcctcacagagcgagcctcc 540
OY 985 cgcgcgcgagggagtttgtgatgcaggtgaaggttggaaccacacactgcagaagaacgg 1044
Db 541 cgcgcgcgagggagtttgtgatgcaggtgaaggttggaaccacacactgcagaagaacgg 600
OY 1045 gcaccaacaagaaggtgcccgaagcgcaatgcagccgagaacatgtggagatccttggt 1104
Db 601 gcaccaacaagaaggtgcccgaagcgcaatgcagccgagaacatgtggagatccttggt 660
OY 1105 tcaaaagtcgccgacgcgccagcccacaaaccgcactcaagtcagaggagaacaccca 1164
Db 661 tcaaaagtcgccgacgcgccagcccacaaaccgcactcaagtcagaggagaacaccca 720
OY 1165 taagaagaacaggggatggaagaaacactttttgaacctggtctggtgatgaa 1224
Db 721 taagaagaacaggggatggaagaaacactttttgaacctggtctggtgatgaa 780
OY 1225 atggactagtaataaagagatgagttcagatgccttatctaagtcacagcagctgc 1284
Db 781 atggactagtaataaagagatgagttcagatgccttatctaagtcacagcagctgc 840
OY 1285 ctgctggaattcttcccatgtgtcccgaggtgcgccagcgtgtaggttagtcaagac 1344
Db 841 ctgctggaattcttcccatgtgtcccgaggtgcgccagcgtgtaggttagtcaagac 900

OY 1345 atcacaccaagaatttaccagggcagctccgaatcctgcgaaagccacgtaactgcc 1404
Db 901 atcacaccaagaatttaccagggcagctccgaatcctgcgaaagccacgtaactgcc 960
OY 1405 tgatagcccgagagtgtgtatgaggggcacctgcgccacag-ccgagaccatttaag 1463
Db 961 tgatagcccgagagtgtgtatgaggggcacctgcgccacagcccgagaccatttaag 1020
OY 1464 aatacatctctcagggccaagctaccccatgacctctcacgagacccttgagcaactg 1523
Db 1021 aatacatctctcagggccaagctaccccatgacctctcacgagacccttgagcaactg 1080
OY 1524 gactatcttccagatccaggatctccaggttgataacaaagactt 1570
Db 1081 gactatcttccagatccaggatctccaggttgataacaaagactt 1127

RESULT 8
X40002
ID X40002 standard; DNA; 946 BP.

XX X40002;
AC XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.

KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.

XX Homo sapiens.

XX WO904265-A2.

XX 28-JAN-1999.

XX 15-JUL-1998; 98WO-US14679.

XX 22-JUN-1998; 98US-0102322.

XX 17-JUL-1997; 97US-0896164.

XX 10-OCT-1997; 97US-0061599.

XX 10-OCT-1997; 97US-0061765.

XX 10-OCT-1997; 97US-0948705.

XX 11-OCT-1997; 97GB-0021697.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;

XX Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;

XX Tureci O;

XX WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

XX isolated using sera from cancer patients, used to develop products

XX for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 629-630; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC Lung cancer.

SQ sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other;

Query Match	25.48;	Score 798.8;	DB 20;	Length 946;
Best Local Similarity	93.68;	Pred. No. 1.2e-192;		
Matches 888; Conservative	0;	Mismatches 54;	Indels 7;	Gaps 6;

QY	445	gagggacagcaatttcaatgycaaaaggaaagacaagacaggtctgcgaaacacgatctgctg	504
Db	1	gaggacagcaatttcaatgycaaaaggaaagacaagacaggtctgcgaaacacgatctgctg	60
QY	505	ccaaagcgttgaggaatcctgcagaaatgagccctccagagaggtctggaagtgaatgaa	564
Db	61	ccaaagcgttgaggaatcctgcagaaatgagccctccagagaggtctggaagtgaatgaa	120
QY	565	gagaaatccgaagaagaanaatctcaataatctgaataaagltcaagtgtttgagatgac	624
Db	121	gagaaatccgaagaagaanaatctcaataatctgaataaagltcaagtgtttgagatgac	180
QY	625	ttaacgcgaacttgcctgtgaatttcgaggtgcccgggaggtgccccaccacatga	684
Db	181	ttaacgcgaacttgcctgtgaatttcgaggtgcccgggaggtgccccaccacatga	240
QY	685	agaacttgtgacccaagtttcggttgggagtttgytggggagagtggaagggaaaaagca	744
Db	241	agaacttgtgacccaagtttcggttgggagtttgytggggagagtggaagggaaaaagca	300
QY	745	agaagatttcaaaagaaaaatgcgcgcatagtctgttcttgaggagctgaaagaattaccgc	804
Db	301	agaagatttcaaaagaaaaatgcgcgcatagtctgttcttgaggagctgaaagaattaccgc	360
QY	805	ccctgcctgcagttgaaacgagtaaaagcctagaatcaaaaaagaaaaaaaaccatagtca	864
Db	361	ccctgcctgcagttgaaacgagtaaaagcctagaatcaaaaaagaaaaaaaaccatagtca	420
QY	865	agccacagacaagcccgaaatatygcaggggatcaalccgaattagccgactgcccaga	924
Db	421	agccacagacaagcccgaaatatygcaggggatcaalccgaattagccgactgcccaga	480
QY	925	tccagcaggcacaaaagagaaagagcagcagatgacacgtctcctcacagagcgagcctcc	984
Db	481	tccagcaggcacaaaagagaaagagcagcagatgacacgtctcctcacagagcgagcctnc	540
QY	985	cgcgccgcagggagtttgtatgcaggtgaaagttgaaaccacac-tgcagaagyaagc	1043
Db	541	cgcgccgcagggagtttgtatgcaggtgaaagttgaaaccacac-tgcagaagyaagc	600
QY	1044	ggcaccacaagaaggttggccaagcgaatgcagccgagaacaatgctgagatccttgt	1103
Db	601	ggcaccacaagaaggttggccaagcgaatgcagcccgagaacaatgctgagatccttgt	660
QY	1104	ttc-aaagttccgcgcgcgcagccaccaccaaccgcactcaagtcagagagaagaacacc	1162
Db	661	ttcaaaagttccgcgcgcgcagccaccaccaaccgcactnaagtcagangagaagaacccc	720
QY	1163	cataaagaacacagggatgtygaagaanaagtaacctttttgaacctggtcttgaggatga	1222
Db	721	altaagaagaacacagggatgtygaagaanaagtlanc--ntttgaanctgctnttggatla	778
QY	1223	aaatgagactagtaataaagagatgagttlcagatgcttatctaagtcatacagcagct	1282
Db	779	aaatgag-ctgtlanlaaagagatgagttlcagatgctntatctaagtcatnaaca-ct	836
QY	1283	gcctgctggaattcttcccatggtgcccgcaggtcgcgccaggtctgtagagttatgcaag	1342
Db	837	tgctgtggaatcttcccatggtggggcgcaggtcncgccagctttagagatla-tcang	895
QY	1343	acatcacacccaagaatttaccaggggcagctccgaatcctgccaagcc 1391	
Db	896	ccnttcncccaagaatttccgcgggcagtttccaatctgccaagccc 944	

RESULT	9
X40000	
ID	X40000 standard; DNA; 844 BP.
XX	
AC	X40000;
XX	
DT	02-JUL-1999 (first entry)

	Query Match	24.0%;	Score 755.6;	DB 20;	Length 844;	
	Best Local Similarity	96.3%;	Pred. No. 1e-181;			
	Matches 812;	Conservative 0;	Mismatches 26;	Indels 5;	Gaps 4;	
QY	607 aagtggttgagatgcaactaaacggaactgcccgtgaatttcagagtggtgcccggaga	666				
Db	1 aagtggttgagatgcaactaaacggaactgcccgtgaatttcagagtggtgcccggaga	60				
QY	667 gtggtccacccccacatgaagaactttgtgaccaaagtttcggttggtggagtttgtgggg	726				
Db	61 gtggtccacccccacatgaagaactttgtgaccaaagtttcggttggtggagtttgtgggg	120				

QY 727 aaggtgaagggaagcaagaagatttcaaagaataatgcgcgcatagctgttctgagg 786
|||
Db 121 aagtggaagggaagcaagaagatttcaaagaataatgcgcgcatagctgttctgagg 180
QY 787 agctgaagaaqgttacccgccccctgctgcagttgtaacgagtaaaagcctagaaatcaaaaaga 846
|||
Db 181 agctgaagaaqgttacccgccccctgctgcagttgtaacgagtaaaagcctagaaatcaaaaaga 240
QY 847 aaacaaaccccatagtcgaagcccaagacaagcccaagaatatgcccagggatcaatccga 906
|||
Db 241 aaacaaaccccatagtcgaagcccaagacaagcccaagaatatgcccagggatcaatccga 300
QY 907 ttagccgactggcccagatcccaagcagcaaaaaaaggaagagcagagtaacagctcc 966
|||
Db 301 ttagccgactggcccagatcccaagcagcaaaaaaaggaagagcagagtaacagctcc 360
QY 967 tcacagagcagggcctcccgccgccaaggagttgtgatgcaggtgaagttggaacc 1026
|||
Db 361 tcacagagcagggcctcccgccgccaaggagttgtgatgcaggtgaagttggaacc 420
QY 1027 acactgcagaaggaacgagcaccacaagaaggtgccaagcgcaatgcagccgagaaca 1086
|||
Db 421 acactgcagaaggaacgagcaccacaagaaggtgccaagcgcaatgcagccgagaaca 480
QY 1087 tgcgtgaagatccttgtttcaaagtcgccgagcgccagcccaaacccgcaactcaagt 1146
|||
Db 481 tgcgtgaagatccttgtttcaaagtcgccgagcgccagcccaaacccgcaactcaagt 540
QY 1147 cagaggaagaagacaccataaagaacccaaggagatggaagaaagtaacctttttgaac 1206
|||
Db 541 cagaggaagaagacaccataaagaacccaaggagatggaagaaagtaacctttttgaac 600
QY 1207 ctgctctgggatgtaaaatgagactagtaataaagagatgagttcagatgcctatc 1266
|||
Db 601 ctgctctgggatgtaaaatgagactagtaataaagagatgagttcagatgcctatc 660
QY 1267 taagtcacagcagctgcctgtcgtgaattcttcccatgtgcgccgagtcgcccagctg 1326
|||
Db 661 taagtcacagcagctgcctgtcgtgaattcttcccatgtgtgcccgangtcgcccaagctg 720
QY 1327 taag-aagtcagtcgaagacatcacacccaaagatttaccagggcagctccgaatcctgccc 1385
|||
Db 721 taggaagttagtcgaagacatnacaccccaagatttaccga-ggcagcttcgaatcctgccc 779
QY 1386 aagggcagcgttaactgcatgatagcccgagagtggtgtatgggggcacctgcgccaca 1445
|||
Db 780 anggcncng--tactgcgatgatagccc-anagltgtgtatgtgggcancttgcgccca 836
QY 1446 gcc 1448
|||
Db 837 ggc 839

RESULT 10
X39992
ID X39992 standard; DNA; 773 BP.
XX
AC X39992;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.

XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundsschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

Query Match 24.0%; Score 753.2; DB 20; Length 773;
Best Local Similarity 99.18; Pred. No. 3.9e-181;
Matches 766; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 799 taccgccccctgctgcagttgaacgagtaaaagcctagaaatcaaaaagaacaaaccca 858
|||
Db 1 taccgccccctgctgcagttgaacgagtaaaagcctagaaatcaaaaagaacaaaccca 60
QY 859 tagtcaagccacagacaagcccaagaatatgcccagggatcaatccgattagccgactgg 918
|||
Db 61 tagtcaagccacagacaagcccaagaatatgcccagggatcaatccgattagccgactgg 120
QY 919 cccagatccagcagcaaaaaaggaagagagccagagatcacgctcctcacagagcgag 978
|||
Db 121 cccagatccagcagcaaaaaaggaagagagccagagatcacgctcctcacagagcgag 180
QY 979 gcctcccgccgcccagggagttgtgatgcaggtgaaggttggaaccacactgcagaag 1038
|||
Db 181 gcctcccgccgcccagggagttgtgatgcaggtgaaggttggaaccacactgcagaag 240
QY 1039 gaacggcaccacaagaaggtgcccacagcgaatgcagccgagaacatgtctgagatcc 1098
|||
Db 241 gaacggcaccacaagaaggtgcccacagcgaatgcagccgagaacatgtctgagatcc 300
QY 1099 ttggttcaaagtcgccgagcgagcccaaaaacccgcaactcaagtcaagggagaaga 1158
|||
Db 301 ttggttcaaagtcgccgagcgagcccaaaaacccgcaactcaagtcaagggagaaga 360
QY 1159 caaccataaagaacccagggagatggaagaaagtaacctttttgaaacctggtctggg 1218
|||
Db 361 caaccataaagaacccagggagatggaagaaagtaacctttttgaaacctggtctggg 420

KW Cancer associated antigen; diagnosis; research; treatment; human;
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer; ss.
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 794 BP; 273 A; 141 C; 152 G; 206 T; 22 other;

Query Match 16.0%; Score 501.4; DB 20; Length 794;
Best Local Similarity 96.3%; Pred. No. 2.4e-117;
Matches 517; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

QY 2598 tcttgacttcaaaaacttgataagctccc--acagttgtataaattggacaattt 2655
DB 718 TCCTGGACTTCAAAACCTTGAATAAGTCTCCCNCCAGTTGNATAAATTGGCCNATTT 659
QY 2656 aggaattttaacttgaatgatcatcttggttccatttttattcattttttgtt 2715
DB 658 AGCAATTTTAAACNNTAGATGTTCATTTGGTTCATTTTATTTCATTTTATTTTGGT 599
QY 2716 aatgcaaacaggaactaaatgaacttgcctctgttttaagattataaaaaacattg 2775
DB 598 AATGCAAAACAGGNCCTTAATGAACCTTGATCTCTGTTTAAAGATTATTAAACAAATTG 539
QY 2776 tctatctatacatatgctcttgaggacttagcttccactacactacagatatgctc 2835
DB 538 TGNATCTATACATATGGCTCTTGAGGACTTAGCTTTCACCTACACTACAGCATATGATCTC 479
QY 2836 catgtagcatataataaactgcagagtgatttccagagtgctcgatactgttaattaca 2895
|||||

DB 478 CATGTAGTCCATATAAACCTGCAGAGTGATTTCCANAGTGCCTCGATACTGTTAATTACA 419
QY 2896 tctccattaggcgctgaagaatgacctaagcttctgtatacagctgtgttcttgat 2955
DB 418 TCTCCNNTTAGGGCTGAAGAAGATGACCTACGTTTCTGTATACANCNTGTGTCTTTGAT 359
QY 2956 gttgtgttactgtacacagaagtgtgtgcactgagcctcgtcgtgtgtccgtatgaaa 3015
DB 358 GTTGTGNTACTGNACACAGAAGTGTGTGCNCTGAGGCTCTGCGTGTGTCGTATGAAA 299
QY 3016 accctgtagccctgcgagftaagtactgcttccattcattgttlaagcttgaatttct 3075
DB 298 GCCTGTAGCCCTGCGAGTTAAGTACTGCTTCATTCATTGTTACGCTGGAATTTTCT 239
QY 3076 ccccatggaatgtaagtaaaacttaagtgttgcatacataaatggttaactaaa 3132
DB 238 CCCCATGGAATGTAAGTAAACTTAAGTGTGTTCATCAATNAATGGTATACTAA 182
|||||

RESULT 13
279981/c
ID 279981 standard; cDNA; 660 BP.
XX
AC 279981;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytosstatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer .
XX
PS Claim 1; Page 158; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other;

Query Match 13.8%; Score 433.8; DB 21; Length 660;

XX WO9964576-A2.
PN
XX 16-DEC-1999.
PD
XX 09-JUN-1999; 99WO-IB01062.
PF
XX 10-JUN-1998; 98US-0088801.
PR
XX (FARB) BAYER CORP.
PA
XX Endege WO, Steinhmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI; 2000-087220/07.
DR
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
XX
PS Claim 15; Page 444; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
XX hyperplasia.
XX
Sequence 708 BP; 226 A; 127 C; 143 G; 191 T; 21 other;

Query Match	10.4%;	Score 327.8;	DB 21;	Length 708;
Best Local Similarity	82.8%;	Pred. No. 2.1e-73;		
Matches 515; Conservative	0;	Mismatches 83;	Indels 24;	Gaps 12;

[illegible][illegible]

Search completed: April 5, 2001, 08:28:38
Job time: 59162 sec

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Herpes simplex virus
STRAIN: Herpes Simplex Virus Type 1
US-08-458-568A-11

Query Match 1.7%; Score 52; DB 1; Length 12001;
Best Local Similarity 56.4%; Pred. No. 0.00063;
Matches 97; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 56 cgcgtctctcgcgtccgcgtctccttgaccgcgtcccccgcgcgcgcgcgc 115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1321 CGCTCTCTCCCCCGCTCTCCCGCTCTCTCCCGCTCTCTCTCTCTCTCC 1380
QY 116 gcctctccacggcactcgcgtctcctcctcctcctcctcctcctcctt 175
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1381 CGCTCTCTCCCCCGCTCTCCCGCTCTCTCCCGCTCTCTCTCTCTCTCC 1440
QY 176 tcctctctcctcctcctcctcctcctcgcgcgcacccagacgcgcgcg 227
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1441 CGCTCTCTCCCCCGCTCTCCCGCTCTCTCCCGCTCTCTCTCTCTCTCC 1492

RESULT 4

US-08-658-136-2
Sequence 2, Application US/08658136
Patent No. 6071717

GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 53526 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match 1.6%; Score 51.2; DB 3; Length 53526;
Best Local Similarity 61.0%; Pred. No. 0.0025;
Matches 100; Conservative 0; Mismatches 63; Indels 1; Gaps 1;

QY 56 cgcgtctctcgcgtccgcgtctccttgaccgcgtcccccgcgcgcgcgc 115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34559 CTCCTCCCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 34618
QY 116 gcctctccacggcactcgcgtctcctcctcctcctcctcctcctcctt 174
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34619 TCCTCTCTCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 34678
QY 175 tcctctctcctcctcctcctcctcctcgcgcgcacccagacgcgc 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34679 CCCCTCTCTCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 34722

RESULT 5

US-08-658-136-1
Sequence 1, Application US/08658136
Patent No. 6071717

GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 53577 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

Query Match 1.6%; Score 51.2; DB 3; Length 53577;
Best Local Similarity 61.0%; Pred. No. 0.0025;
Matches 100; Conservative 0; Mismatches 63; Indels 1; Gaps 1;

QY 56 cgcgtctctcgcgtccgcgtctccttgaccgcgtcccccgcgcgcgcgc 115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34554 CTCCTCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 34613
QY 116 gcctctccacggcactcgcgtctcctcctcctcctcctcctcctcctt 174

[illegible]

RESULT 6
US-09-128-155-16
; Sequence 16, Application US/09128155

```

; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

Query Match	1.5%;	Score 47.6;	DB 3;	Length 152331;
Best Local Similarity	51.7%;	Pred. No. 0.046;		
Matches 107; Conservative	0;	Mismatches 100;	Indels 0;	Gaps 0;

QY	56	cgcgtctctcgcgctcccgcttcccttgacgcgcctcccccccggccgcgcgcc	115
Db	21895	ccccccccctccccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	21954
QY	116	gcctctccaagcgaactcgcgcctcttcctctccttcgttcctttctctctctttt	175
Db	21955	ccccgc	22014
QY	176	tcccttctcttcccccttccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	235
Db	22015	ccccccccccgc	22074
QY	236	ggaagcagcagccagaagaataacccc	262
Db	22075	gcaccccccaaccccgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	22101

RESULT 7
5196516-7
Patent No. 5196516
APPLICANT: SCHREURS, CHRISTA S.;METTENLEITER, THOMAS C
; SIMON, ARTUR J.;LUKAS, NOEMI;ZIHHA, HANNS J.
TITLE OF INVENTION: PSEUDORABIES VIRUS VACCINE
NUMBER OF SEQUENCES: 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/383, 833
FILING DATE: 21-JUL-1989
SEQ ID NO:7:
LENGTH: 4897
5196516-7

Query match 1.48; Score 45; DB 5; Length 4897;

Best Local Similarity	54.58;	Pred. No. 0.031;
Matches 90; Conservative	0;	Mismatches 75; Indels 0; Gaps 0;

QY 53 ctgcgcgtctctctcgcgtccgcgtctctcttgaacgcgcctccccccccgcggccgcgcgcg 112
 || || |||| | |||| || || | || |||| || || |
 Db 4347 ccgtccctctcccccgcgtccctctcccccgcgtccctctcccccgcgtccctctccct 4406
 QY 113 ccgcgcctctcccaagcgcactccgcgcctctctccctccctctcgtccctctctccctctc 172
 || || |||| | | | | || |||| || | |||| |||| |
 Db 4407 ccgtccctctcccccgcgtccctctctcccccgcgtccctctcccccgcgtccctctccct 4466
 QY 173 ttctcctctctccttccctctccctctcgcgcgcacccgcacggacgcg 217
 ||| ||| || |||| |||| || | | ||| |||| |
 Db 4467 ccgtccctctcccccgcgtccctctctcccccgcgtccgcgcgacccac 4511

RESULT 8
US-08-455-550-7/c
; Sequence 7, Application US/08455550
; Patent No. 5670338

```

: GENERAL INFORMATION:
: APPLICANT: MURAKAMI, KAZUO
: APPLICANT: UENO, NAOTO
: APPLICANT: KATO, YUKIO
: TITLE OF INVENTION: XENOPUS LAEVIS BONE MORPHOGENETIC PROTEINS AND USE THE
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dike, Bronstein, Roberts & Cushman
: STREET: 130 Water Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
:

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,550
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/056,564
; FILING DATE: 30-APR-1993
; APPLICATION NUMBER: 07/577,892
; FILING DATE: 05-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Eisenstein, Ronald I
; REGISTRATION NUMBER: 30628
; REFERENCE/DOCKET NUMBER: 40302-FWC-DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: 200291
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1558 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
;
US-08-455-550-7

```

Query Match	1.4%;	Score 44.6;	DB 1;	Length 1558;
Best Local Similarity	58.8%;	Pred. No. 0.02;		
Matches 77; Conservative	0;	Mismatches 54;	Indels 0;	Gaps 0;


```

: APPLICANT: Schaffer, Priscilla A.
: APPLICANT: Yeh, Lily
: TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
: TITLE OF INVENTION: Infections
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,568A
: FILING DATE: 02-JUNE-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/065,146
: FILING DATE: 05-MAY-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary Ph.D., Kathryn R.
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: DFCI-0029
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12001 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Herpes simplex virus
: STRAIN: Herpes Simplex Virus Type 1
: US-08-458-568A-11

Query Match          1.4%; Score 44.2; DB 1; Length 12001;
Best Local Similarity 58.9%; Pred. No. 0.086;
Matches 76; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 90 ctcccccccgccgagcgagccgctctccagcgcaactcgcctctccctccc 149
   1 111 11111 111 111 111 1 1 1 1 11111 111 111
Db 7697 cccccctccgccccccgctccctccgccccccgctccctccgcccc 7638

OY 150 ttctgctctctctctctctctctctctctctctctctctctctctctctc 209
   111111 11 1 1111 111 11111 11 111 11 11 11 11 11
Db 7637 CGCGTCCCTCTCTCGCGCCCGCGGTCCCGCTCTCTCTCTCTCTCTCTCT 7578

OY 210 aggacggcc 218
   1 11 11
Db 7577 CCGCCACCC 7569

RESULT 12
US-08-323-443B-1
: Sequence 1, Application US/08323443B
: Patent No. 5654170
: GENERAL INFORMATION:
: APPLICANT: KLINGER, KATHERINE W.
: APPLICANT: LANDES, GREGORY M.
: APPLICANT: BURN, TIMOTHY C.
: APPLICANT: CONNORS, TIMOTHY D.
```

```

: APPLICANT: DACKOWSKI, WILLIAM R.
: APPLICANT: GERMINO, GREGORY
: APPLICANT: QIAN, FENG
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Darby & Darby PC
: STREET: 805 Third Avenue
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/323,443B
: FILING DATE: 12-OCT-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Ludwig, S. Peter
: REGISTRATION NUMBER: 25,351
: REFERENCE/DOCKET NUMBER: 0372/0A462
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 527-7700
: TELEFAX: (212) 753-6237
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 31571 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: PKD1 GENOMIC
: US-08-323-443B-1

Query Match          1.3%; Score 41.4; DB 1; Length 31571;
Best Local Similarity 53.1%; Pred. No. 0.9;
Matches 111; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

OY 1 acttcctgcccgggtgagcgagcgtctcagcgtttgcgagcgagcgtcgcg 60
   111 11 111111 111 111 111 1 111 11 11111 1
Db 9941 ACTCCCGAGGGGCTGAGGGGCGCT--GCGCCAGATGACAGCTTGCTGGTGCACAGC 9998

OY 61 ctctctcggtcccgctctctttgaccgcctccccccccggccggcgagccgc 120
   1 111 1 111 111 111 1 111111 11 1 1 11 11 1
Db 9999 CCCTCCACCTCTCCCTCCCTGCGACGCCCTCCACACTCTCCCTCCCTGCGACGCCCTCC 10058

OY 121 ctccaagggcaactcgcgctctccctccctcgttcccttctctctccctttttcctt 180
   1 11 1 11 1 111 11 11 11111 1 11 11 11 11 111
Db 10059 CACCTCTCCCTCCCTGCGACGCCCTCCACACTCTCCCTCCCTGCGACGCCCTCTCCACCTC 10118

OY 181 ctctctccctctctcgcgcgcacggccc 209
   1 1 1 11111 11
Db 10119 TCCCTCCCTGCGACGCCCTCCACACTCTC 10147

RESULT 13
US-08-781-891-79
: Sequence 79, Application US/08781891
: Patent No. 6090620
: GENERAL INFORMATION:
: APPLICANT: Fu, Ying-Hui
: APPLICANT: Yu, Chang-En
: APPLICANT: Oshima, Junko
: APPLICANT: Mulligan, John T.
```

APPLICANT: Schellenberg, Gerald D.
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 TITLE OF INVENTION: WERNER'S SYNDROME
 NUMBER OF SEQUENCES: 209
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,891
 FILING DATE: 27-DEC-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6090620tenburg Ph.D., Carol
 REGISTRATION NUMBER: 39,317
 REFERENCE/DOCKET NUMBER: 240052.419
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 79:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 87350 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

Query Match	1.38;	Score 41;	DB 3;	length 87350;
Best Local Similarity	56.28;	Pred. No. 2.1;		
Matches 77; Conservative	0;	Mismatches 60;	Indels 0;	Gaps 0;

QY	62	tcctcgcgcctccgccttcctcttgacgcctcccccccgccgcgcgcgcctcc	121
Db	55222	TCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC	55281
QY	122	tcacagccacatccgcctcttcctccctctgtccctctctctccctcttcttccttc	181
Db	55282	TTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC	55341
QY	182	ttccttcctccctccctgc	198
Db	55342	TTCTTTCTCTCTCTCTC	55358

```

RESULT 14
US-08-147-777-3
; Sequence 3, Application US/0814777
; Patent No. 5914265
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN R1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
;

```

```

1 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
2
3 COMPUTER: IBM PC compatible
4
5 OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
6
7 SOFTWARE: Wordperfect (Version 5.1)
8
9 CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/08/147,777
11
12 FILING DATE:
13
14 CLASSIFICATION: 800
15
16 PRIOR APPLICATION DATA:
17 PRIOR APPLICATION DATA: including application
18 PRIOR APPLICATION DATA: described below:
19
20 APPLICATION NUMBER: 07/876,289
21
22 FILING DATE: April 30, 1992
23
24 APPLICATION NUMBER: Unassigned (204/144)
25
26 FILING DATE: October 29, 1993
27
28 ATTORNEY/AGENT INFORMATION:
29
30 NAME: Warburg, Richard J.
31
32 REGISTRATION NUMBER: 32,327
33
34 REFERENCE/DOCKET NUMBER: 204/153
35
36 TELECOMMUNICATION INFORMATION:
37
38 TELEPHONE: (213) 489-1600
39
40 TELEFAX: (213) 955-0440
41
42 TELEX: 67-3510
43
44 INFORMATION FOR SEQ ID NO: 3:
45
46 SEQUENCE CHARACTERISTICS:
47
48 LENGTH: 24979 base pairs
49
50 TYPE: nucleic acid
51
52 STRANDEDNESS: single
53
54 TOPOLOGY: linear
55
56 MOLECULE TYPE: DNA (genomic)
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
```

Query Match	1.38;	Score 40.8;	DB 2;	Length 24979;
Best Local Similarity	59.7%;	Pred. No. 1.1;		
Matches	86;	Conservative	0;	Mismatches 57;
			Indels	1;
			Gaps	1;

[illegible]

```

RESULT 15
US-08-452-872-3
; Sequence 3, Application US/08452872
; Patent No. 6057298
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:29:19 ; Search time 11371 Seconds
(without alignments)
1982.509 Million cell updates/sec

Title: US-09-316-048-3
Perfect score: 3217
Sequence: 1 acttcctgcggtcgcggg.....taatactaaaaaaaaaaaaa 3217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp14:*
121: em_estp15:*
122: em_estp16:*
123: em_estp17:*
124: em_estp18:*
125: em_estp19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: gb_gss7:*
159: gb_gss8:*
160: gb_gss9:*
161: em_gss5:*
162: em_gss6:*
163: em_gss7:*
164: em_gss8:*
165: em_gss9:*
166: em_gss10:*
167: em_gss11:*
168: gb_gss10:*
169: gb_gss11:*
170: em_gss12:*
171: gb_gss12:*
172: gb_gss13:*
173: gb_gss14:*
174: gb_gss15:*
175: gb_gss16:*
176: gb_gss17:*
177: gb_gss18:*
178: gb_gss19:*
179: em_gss13:*
180: gb_gss20:*
181: gb_gss21:*
182: gb_gss22:*
183: gb_gss23:*
184: gb_gss24:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1757.8	54.6	1801	146	U69197	U69197 U69197 Soar
C 2	743.4	23.1	829	106	BE300155	BE300155 600944602
	701	21.8	1018	136	BE867390	BE867390 601442470
						BE299828 600944602
4	684.8	21.3	831	106	BE299828	BE299828 600944602
5	646.2	20.1	1036	137	BE889880	BE889880 601512132
6	625.6	19.4	686	40	AW160378	AW160378 au72c03.y
7	621.2	19.3	724	97	AW952360	AW952360 EST364430
8	617.2	19.2	752	137	BE894295	BE894295 601437666
C 9	606.4	18.8	821	107	BE379687	BE379687 601159257
	604	18.8	609	107	BE380123	BE380123 601159257
			686	110	BE613972	BE613972 601504093
10	603.8	18.7	674	38	AV707406	AV707406 AV707406
11	602.2	18.7	616	111	BE674157	BE674157 7d76a04.x
12	594.4	18.5	594	40	AW162002	AW162002 au72c03.x
13	588.2	18.3	669	28	AL037151	AL037151 DKF2p5640
C 14	576	17.9	612	38	AV695897	AV695897 AV695897
15	571.4	17.8	749	110	BE617223	BE617223 601441906
16	565.4	17.6	653	87	AW163731	AW163731 au97f05.y
17	562.2	17.5	609	136	BE810330	BE810330 MR4-PT005
18	555	17.3	650	134	BE082712	BE082712 RC2-BT064
19	551.6	17.1	550	107	BE348298	BE348298 hw16b01.x
C 20	550	17.1	564	40	AW163206	AW163206 au97f05.x
21	548	17.0	584	136	BE813737	BE813737 RC1-BN003
C 22	542.4	16.9	673	135	BE781223	BE781223 601469159
23	534	16.6	573	3	AA206573	AA206573 zq51f08.r
24	532.2	16.5	536	136	BE837505	BE837505 RC2-FN009
C 25	531.2	16.5	866	136	BE868328	BE868328 601443887
26	531.2	16.5	910	138	BE966164	BE966164 601660093
27	517.8	16.1	543	134	BE089861	BE089861 RC5-BT070
C 28	513	15.9	513	13	AA889669	AA889669 ak54a04.s
29	511.4	15.9	512	27	AI983007	AI983007 wt46902.x
C 30	510.4	15.9	852	137	BE872137	BE872137 601446238
31	508.8	15.8	673	109	BE541462	BE541462 601067913
C 32	500.6	15.6	633	89	AW320695	AW320695 uc22b10.y
33	497.4	15.5	569	3	AA191622	AA191622 zp81g10.s
34	488.2	15.2	516	25	AI819766	AI819766 wj42a10.x
C 35	487	15.1	485	25	AI802592	AI802592 wf16a03.x
C 36	483.4	15.0	665	106	BE278193	BE278193 601180017
37	481	14.9	872	135	BE779672	BE779672 601464479
38	480.2	14.9	480	134	BE042598	BE042598 ho26e03.x
C 39	478.4	14.9	494	89	AW351909	AW351909 RC0-CT020
C 40	475	14.8	509	142	N31181	N31181 yx64f02.r1
41	470.4	14.6	517	7	AA427366	AA427366 zw33c05.r
42	470	14.6	517	92	AW580332	AW580332 IL2-HT044
C 43	470	14.6	703	137	BE888420	BE888420 601514177
44	467.4	14.5				
45						

ALIGNMENTS

RESULT 1
LOCUS U69197 1801 bp mRNA
DEFINITION U69197 Soares infant brain IN1B Homo sapiens CDNA clone 22368, mRNA
ACCESSION U69197
VERSION U69197.1 GI:2739420
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1801)
AUTHORS Volorio,S., Simon,G., Repetto,M., Cucciardi,M., Banfi,S., Borsani,G., Ballabio,A. and Zollo,M.
TITLE Sequencing analysis of forty-eight human image cDNA clones similar to drosophila mutant protein
JOURNAL DNA Seq. 9 (5-6), 307-315 (1998)
MEDLINE 99452388
COMMENT Contact: Zollo, Massimo
Telethon Institute of Genetics and Medicine
Via Olgettina 58, Milan, MI 20132, Italy
Email: zollo@tigem.it.
FEATURES
Source location/Qualifiers
1.1801
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p13.1"
/clone="22368"
/clone_lib="Soares infant brain 1N1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lactimid BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGAGAATTCGGCGCCGACAGAAATTTTATTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lactimid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 503 a 409 c 386 g 503 t
ORIGIN

Query Match 54.6%; Score 1757.8; DB 146; Length 1801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 1446 gctccgaatcctgcgaagcgtaactgcatatagcccgagagttgtgtatgg 1505
|||
DB 5 GCTCCGAATCTGCCAAGGCCACGGTAAGTGCATGATAGCCCGAGAGTTGTTGATGGG 64
OY 1506 ggcacctgcgcacagccagaccatttaagaataacatctctcaggccagtaacc 1565
|||
DB 65 GGCACCTCGCCACAGCCGAGACCATTTTAAAGAAATACATCTCTTCAGGCCACGTACC 124
OY 1566 catgacctctcacgagacctctgagcaactgactatcttccagagtccaaggatc 1625
|||
DB 125 CATGGACCTCTCAGAGACCCCTCGAGCAACTGAGCTATCTTCCAGAGTCCAGGATTC 184
OY 1626 caggttgaatacaaaagacttcccaaaaaacaacaagaacgaatttgtatctctatcaat 1685
|||
DB 185 CAGGTTGAATACAAAGACTTCCCAAAAACAACAAGCAAGAAATTGTATCTTTATCAAT 244
OY 1686 tgcctctcagccacctctgatacagcatggtatcgcaaggatgtgagtlcctgcat 1745
|||
DB 245 TGCTCTCTCAGCCACCTCTGATCAGCCATGGTATCGCAAGGATGTGGAGTCTGCCAT 304
OY 1746 gatatgctgcgctgaacatcttaagttgtctgtcagttgagccaacaagtaacagag 1805
|||
DB 305 GATATGGCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGGACCAACAAGTACAGAG 364
OY 1806 atgccaagaacaggaacggaacatgtctgtgtggaaggtgtgaaccttctcggc 1865
|||
DB 365 ATGCCAAGAACAGAAACGAGCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGC 424
OY 1866 catgaacatataaaatcccaacatatatactgaanaaactgaaactgcttggaaat 1925
|||
DB 425 CATGAACCATTTAAATCCCAACATATATACTGAATAACTGAAACTGCTTTGAAATTT 484

OY 1926 tggaaattctgtatacctccagctggyccgagagacacggttggtaagatgtggyccagca 1985
|||
DB 485 TGGAAATTCTGATACCTCCAGTGGGCCGAGAGACACCGGTGGGTAAGGATGTGGGCAGCA 544
OY 1986 gcagggaagacaacagaaacacaaggaaggcgctgtgccc-ggctggaactgtctgggt 2044
|||
DB 545 GCAGGGAAGACAACAGAAACACAGAGAGCGGCTGTGGCCGGGCTGACTGTCTGGGGT 604
OY 2045 ttgtgtgatgcccactcgtgtgacctgtgcygttcctcgaatagcagctgctgtggg 2104
|||
DB 605 TTGTTGTGATGGCCACTCGGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCTGTGGGG 664
OY 2105 aagaaggctgcccagccagctgtgtctcccgagacaccagcagatccacacctgggca 2164
|||
DB 665 AAGAAAGGCTGCCCCAGCCAGCTGGTTCTCCCGGAGACACAGCATCCACACCTGGGCA 724
OY 2165 cctcgtgtttgtgtcttttttccctgtgtgaagaagaacgycagcagccctctc 2224
|||
DB 725 CCTCCGTGTTGGTCTTTTTCCTCCCTGTGTGAAGAAGCAAAACGGCAGACCCCTTCTC 784
OY 2225 aagctgctcactcagacacatlyggacaacaccctgacagccatgcccagagaaggcct 2284
|||
DB 785 AAGCTGCTCACTCAGACACATTGGGACAACCCCTGGACGCCATGCCAGAGAGGCGCT 844
OY 2285 ttgaccgccccagagctaaagcacccagagaanaatcaatgcttcctactcagcgtgac 2344
|||
DB 845 TTGACCGGCCCCAGAGCTAAAGAGCACAGAGAAATCAATGCTTCTACTACGCGTGAC 904
OY 2345 ccaactttctagtgtgccaagccccaccaacctctgtgaagtaaccacacatcaacct 2404
|||
DB 905 CCAACTTTCTAGTGTGCCAGCGGCCCAACCACTCCTGCAGTACCCACACCATCACCCT 964
OY 2405 gcttctcttccaacagagtgaatctgtattcttagtttcattatttcttltgatatg 2464
|||
DB 965 GCTTCTCTTCCAACAGTGATCTGTATCTTAGTTTCATTTTCTTGTGATGATATG 1024
OY 2465 acactataaaaaatttcaattgagaatttctcaatgtatctagttaaatagcacaglt 2524
|||
DB 1025 ACACATATATAAATTTTCATTTGGAATTTCTCAATGTATCTAGTTAAATAGCACAGTT 1084
OY 2525 tggaactgtctgagactgacttatacaataatctaacgcgaanaagatcatatccatgt 2584
|||
DB 1085 TGGAAACTGTCTGAGACTGACTTATCAATAATCTAACCCGACAAAGATCATATCCATGT 1144
OY 2585 gtatgtgttagacatttttatttcaattgactaacccagagacgtttcagtgatgcaaat 2644
|||
DB 1145 GTATGTGTTAGACATTTTATTTTCATTTGACTAACCCAGAGCAGTTTCAGTGATGCAAT 1204
OY 2645 tgtgtgccctctgtgtcagctgtaaacagtcctggaacttcaaaaaaaccttgaataagtc 2704
|||
DB 1205 TGTGTGCCCTCTGTGTTCAGCTGAACACAGTCTCGACTTTCAAAAACTTGAATAGTCTC 1264
OY 2705 ccacagtgtataaaatttgacaatttaggaattttaaacttagatgatctgtgtcc 2764
|||
DB 1265 CCACAGTTGTATAAATTGGACAATTAGGAATTTTAAACTTTAGATGATCATTTGGTTCC 1324
OY 2765 attttattcatatttattttgttgaatgcaacagagacttaaatgaacttgatctct 2824
|||
DB 1325 ATTTTATTTTCATTTTATTTTGTGTAATGCAAAACAGACTTAATGAACCTTGATCTCT 1384
OY 2825 gttttaagaattataaaaaaacatgtgtatctatatacatatggtctctgagacttagct 2884
|||
DB 1385 GTTTTAAAGATTATTAATAAACATTTGTGTATCTATACATATGGCTCTTGAGGACTTAGCT 1444
OY 2885 ttcactacactacagatatgatctcccatgtagtcacatataaacctgcagagtgatcttc 2944
|||
DB 1445 TTCACTACACTACAGGATATGATCTCCATGTAGTCCATATTAACCTGCAGAGTGATTTTC 1504
OY 2945 cagagtgtcagatactgttaattacatctccatlagggctgaaaaagaatgaactacgtlt 3004
|||
DB 1505 CAGAGTGTCTGATACCTGTTAATTTACATCTCCATTTAGGGCTGAAAAAAGATGACCTTAGCTTT 1564
OY 3005 ctgtatacagctgtgtgtctttagatgtgttactgtatacagaagtgltgtgcaactga 3064

|||||
Db 1565 CTGTATACAGCTGTGTTGCTTTTGATGTGTGTACTGTACACAGAGTGTGTGCACTGA 1624
QY 3065 ggctctgcgtgtgtccglatatgaaacctgtagccctgcgagtgtaagtactgtctcca 3124
Db 1625 GGCTCTGCGGTGTGGTCCGTATGGAAGCCCTGTAGCCCTGCGAGTTAAGTACTGTCCA 1684
QY 3125 ttcaattgttaacgctggaattttctcccccattggaatgtgaagtaaaacttaagtgttgt 3184
Db 1685 TTCAATTGTTTACGCTGGAATTTTCTCCCCATGGAATGTAAGTAAACTTAAGTGTGTGT 1744
QY 3185 catcaataaatgttaataactaaaaa 3217
Db 1745 CATCAATAAATGTAATACTAAAAA 1777
RESULT 2
LOCUS BE300155/c 829 bp mRNA EST 20-JUL-2000
DEFINITION 600944602T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960593 3',
mRNA sequence.
ACCESSION BE300155
VERSION BE300155.1 GI:9183903
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM54 row: d column: 02
High quality sequence start: 41
High quality sequence stop: 799.
FEATURES
source location/Qualifiers
1. 829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960593"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 190 a 200 c 226 g 213 t
ORIGIN
Query Match 23.1%; Score 743.4; DB 106; Length 829;
Best Local Similarity 98.8%; Pred. No. 2.8e-172;
Matches 770; Conservative 0; Mismatches 6; Indels 3; Gaps 2;

|||||
Db 744 TCGTCCTTCTTCCTCTCCCTTTTTCCTTCCTTCCTTCCTCCCTCCCTCCGCGCCACGCCAG 685
QY 212 gaccgcccggcgaggagcagctcggagcagcagcagagttatttaaccacttaaccctc 271
Db 684 GACCGCCGGCCGGGGAGCAGAGCTCGGAGCAGCAGCAGAGTTATTAACTTAACCTTAACCTC 625
QY 272 tcagaactgaacaagaacaacattgttcctggaacgcccctcttlttaaaaaaagaagcat 331
Db 624 TCAGAACTGAACAAGACAACATTTGTCCTGG-ACGCCCTCTTTTAAAGAAAGCAT 566
QY 332 aaccctactgtagaactaaatgcactgtgcatgaaacttggaaaaaaaccatgtataa 391
Db 565 AACCCCTACTGTAGAACTAAATGCACCTGTGCACTTGAAACTTGAAAAAACCATGTATAA 506
QY 392 gcctgttgacccttactctcgatgcaagtcacactataactacaatgagaggtgc 451
Db 505 GCCTGTGACCCCTTACTCTCGGATGCGAGTCACCTATACTACACATGAGAGAGGTGC 446
QY 452 ttatcccccgaggtacttttaccatttcacagttccacccttacttatacaagtgaact 511
Db 445 TTATCCCCGAGGTACTTTTACCATTTCACAGTTCCACCTTTACTTATCAAGTGAAGT 386
QY 512 ttctgtggaagacgcaatttaatgccaaggaagaagaagaagcgtgcgaacacga 571
Db 385 TTCTGTGGAGGACAGCAATTTAATGGCAAGGAAGACAAGACAGGCTGCGAAACACGA 326
QY 572 tgctgtgccaaagcgttgagatcctgcgaatlgagccctgcagagagcgtgaggt 631
Db 325 TGCTGTGCCAAAGCGTTGAGGATCCTGCAGAAATGAGCCCCCTGCCAGAGAGGCTGAGGT 266
QY 632 gaatgaaagaatccgaagaagaataatcctaataatctgaataagtcagtgttga 691
Db 265 GAATGGAAGAAGAAATCCGAAGAAATCAATAAATCTGAATAAGTCAAGTGTTTGA 206
QY 692 gattgcacttaacggaaacttgccctgtgaatttcgaggtgagtgccgggagagtgcccacc 751
Db 205 GATTGCACCTTAACGGAAGACTTGCCCTGTGAATTTGAGAGTGGCCGGAGAGTGCCAC 146
QY 752 ccacatgaagaactttgtgaccaaggttccggttgaggagtttggggggaagtgaag 811
Db 145 CCACATGAAGAAGCTTTGTGACCAAGGTTTCGTTGGGAGTTTGTGGGGAAGGTGAAG 86
QY 812 gaaaagcaagaagatttcaaaagaaaatgcgcgcataagctgttcttgaggagctgaaga 870
Db 85 GAAAAGCAAGAAGATTTCAAAGAAAAATGCCGCCTAGCTGTTCTGAGAGGAGCTGAAGA 27
RESULT 3
LOCUS BE867390 1018 bp mRNA EST 27-SEP-2000
DEFINITION 601442470F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846889 5',
mRNA sequence.
ACCESSION BE867390
VERSION BE867390.1 GI:10316166
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1018)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

QY 244 agccagagttatlaaccacttaacctctcgaactgacaagaacacattgttcctg 303
|||||
Db 180 AGCCAGAGTTATTAACCACTTAACCTCTCAGAACTGACAAAGACACATGTCTCTGG 239
QY 304 aacgccctcttlttaaaaaagaagcataaccctactgttagaactaaatgcaactgtgca 363
|||||
Db 240 AACGCCCTCTTTTAAAGAAAGCATACCCCTACTGTAGAACTAAATGCACCTGTGCA 299
QY 364 tgaaccttggaaaaaaaccatgtataagcctgttgaccctactcgcgagtcagtcaca 423
|||||
Db 300 TGAACCTTGAAAAAACCAATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCA 359
QY 424 cctataactacaacatgagagagtgcttatcccccgagtgacttaccattccag 483
|||||
Db 360 CCTATACTACACATGAGAGAGGTGCTTAATCCCGAGGTACTTTTACCCATTTCAG 419
QY 484 ttccaccttacttatacaagtgaacttctgttgaggagacagcaattaatgcaag 543
|||||
Db 420 TTCACCTTTACTTTATCAAGTGAACCTTCTGTGGAGAGACACAAATTAAATGCAAAAG 479
QY 544 gaaagacaagacagctgcgaacacagatgctgtgccaaagcgttgagatcctgcaga 603
|||||
Db 480 GAAAGACACAGACAGCTGCCGAACACGATGCTGCTGCCAAAGCGTTGAGGATCTTCACAG 539
QY 604 atgagccctgcccagaagagcgtgagtgtaatggaagaaatccggaagaagaatctca 663
|||||
Db 540 ATGAGCCCCCTGCCAGAGAGCGCTGGAGGTGAATGGAAGAAATCCGAAGAAGAAATCTCA 599
QY 664 ataaatctgaataaagtaagtggttgagactgaactaaacggaacttgctgtgaat 723
|||||
Db 600 ATAAATCTGAATAATAGTCAAGTG-TTGAGATGGCACTTAAACGGAACCTTGCTGTGAAT 658
QY 724 tcgaggtgccccgagagagtgcccaacccacatgaagaacttgaccacaggttcg 783
|||||
Db 659 TCGAGGTGGCCCGGAGAGTGG-CCACCCCATGAGAACTTTGTGACCA--GGTTCCG 715
QY 784 ttggggagttgtgggggaaggtgaaggaaaaagcaagaagattccaagaaaaaatgc 841
|||||
Db 716 GTGGGGAGT--AGGGGGACAGGTGACGGAAGCCAGAGATTCCAAGGAAAAATGC 770

RESULT 5
BE889880 1036 bp mRNA EST 29-SEP-2000
LOCUS 601512132F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3913590 5',
DEFINITION mRNA sequence.
ACCESSION BE889880
VERSION BE889880.1 GI:10347645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1036)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM720 row: p column: 07
High quality sequence stop: 706.
Location/Qualifiers
1. 1036
/organism="Homo sapiens"

FEATURES
Source

/db_xref="taxon:9606"
/clone="IMAGE:3913590"
/clone_id="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 296 a 230 c 220 g 290 t
ORIGIN

Query Match 20.1%; Score 646.2; DB 137; Length 1036;
Best Local Similarity 91.5%; Pred. No. 2.5e-148;
Matches 740; Conservative 0; Mismatches 63; Indels 6; Gaps 5;

QY 2290 cggccccagagctaaagcaccagagaagaatcaaatgcttccctactcagcgtgacccaac 2349
|||||
Db 1 CGGCCCCAGAGCTAAAGCACCAGAGAAATCAATGCTTCTCTACTCAGCGTGACCCAAC 60
QY 2350 ttctctagtgccacggcccccaacctctgcagtlaccacacacataccaactgcttc 2409
|||||
Db 61 TTTCTAGTGTGCCAGGGCCCCACACCTCTCTGACGTACCACACATCACCACTGCTTT 120
QY 2410 ctctccaacagtgatctgtatctcttagttcatatttcttcttgatgatgacact 2469
|||||
Db 121 CTCTTCCAACAGTGAATCTGTATTCTTAGTTCAATTATTTCTTTGATTGATGACACT 180
QY 2470 atataaaatttcatattgagaatttctcaattgtatctagtttaataagcacagtttgaa 2529
|||||
Db 181 ATATAAATTTTCATTTGAGAATTCTCAATTGTATCTAGTTAAATAGCACAGTTTGAA 240
QY 2530 actgtctgagactgaacttatacaataatcaaccgacaagaatcatatccatgtgatg 2589
|||||
Db 241 ACTTGCTGAGACTGACTTTATCAATAATCTAACCGACAAAGATCATATCCATGTGTATG 300
QY 2590 tggtagacatttatttcaaltgaactaaccagagacagtltcagtgatgcaaatgtgt 2649
|||||
Db 301 TGGTTAGACATTTTATTTTCATTGACTAACCCAGGACAGTTTCAGTGTCAAAATGTGT 360
QY 2650 gccctctggttcagctgaacagctcctgacttccaanaaccttgaaataagtcctccaca 2709
|||||
Db 361 GCCCTCTGCTGACCTGAACACAGTCCCTGACTTCAAAAACCTTGATAAAGTCTCCACAC 420
QY 2710 gttgtataaattgacaatttagaattttaacttttagatgatcatgtgtccattt 2769
|||||
Db 421 GTGTATAAATTGACAAATTTAGGAATTTAAACTTAGATGATCATTTGTTCCATT 480
QY 2770 tatttcatttatttctgttaatgcaaacaggaacttaaatgaactttgatctctgtt 2829
|||||
Db 481 TATTTCAATTTTATTTTGTATATGCAACAGGACTTAATGAACTTTGATCTCTGTTT 540
QY 2830 aaagattataaaaaacatgtgtatctatacatatgctcttgaggacttagcttcac 2889
|||||
Db 541 AAAGATTATTAAAAACATTTGTGTATCTATCATATGAGCTCTTGAGGACTTAGCTTTCA 599
QY 2890 taactacagatatgctcc-atgtagtccatataaaccctgcagagtg--atttcca 2946
|||||
Db 600 TACACTACAGATATGATCTCCCAATGTAGTCCATATAAACCTGCAGAGTGTTCACG 659
QY 2947 gagtgtcgtact-ghtaattacatctcatttaggctgaaaaaagaatgaactacgttcc 3005
|||||
Db 660 AAGTGTGATACTGCTTTCATTACTCTCCATTAGGGTGAAAAAGAAATGACCTAGTTCTG 719
QY 3006 tgtatacagctgtgtgtcttcttgatgtgtgttactgtacacagaagtgltg-tgcactga 3064
|||||
Db 720 TATAAAGTGTGTGCTTGACAGGTTGCGGGTTACTGTGAACCAAGAGTGTGTCCCTGA 779
QY 3065 ggcctcgtgtgtgtccgtatgaaaaac 3093
|||||
Db 780 GGGTCTGAGATGTGGCCCGCTGTATACC 808

[illegible]

Db	120	CGCCTCTTCCCTCCCTTCGTCCCTTCTTCTCTCCCTTTTTCCTTCTTCCCTTCCCTCC	179
OY	195	tcgcccacccgcccagagaccgcgcgcgggagcagctcggagcagcagccagagttt	254
Db	180	TCGCCGCACCGCCAGGACCGCCGCGCGGGGACGAGCTCGGAGCAGCCAGAGTTT	239
OY	255	attaaccacttaacctctcagaactgacaaagacacacatgttccctggaacgccctct	314
Db	240	ATTAACTTAACCTCTCAGAACTGAAACAAGACAACATTTGTTCCGTGAACGCCCTCTT	299
OY	315	tttaaaaaagaagcataaacccctactgttagaactaaatgacactgtgtgataaacttga	374
Db	300	TTTAAAAAAGAAAGCATTAACCCCTACTGTAGAACTAAATGCACCTGTGCATGAACCTTGA	359
OY	375	aaaaaaccaatgtataagcctgtltgacccttactctcgatgtcagtcacactataactac	434
Db	360	AAAAAACCAATGTATTAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATAACTAC	419
OY	435	aacatgagagagtgctctatcccccgagtgacttlaaccatttccagttccacctta	494
Db	420	AACATGAGAGGAGGTGCTTATPCCCCGAGGTACTTTTACCATTTCAGTTCCACCTTTA	479
OY	495	cttatcaagtgaactttctgtgagagacagcaattlaatgycaaaggaagaacaaga	554
Db	480	CTTTATCAAGTGAACCTTCTGTGGGAGGACAGCAATTTAATGGCAAGAAAGACAAGA	539
OY	555	cagcctgcgaacacgatagtctgtcccaagcgttgagatccctgcagatgagccctg	614
Db	540	CAGGCTGCANACACGATGCTGCTGCCCAAGCCTTGAGGATCCTGCAGAAATGAGCCCCGTG	599
OY	615	ccagagagcctgaggtgtaatggaagagaatccgaagaagaatatctcaataaatctgaa	674
Db	600	CCAGAAGAGGCTGAGGTGAATGGAGAAGAAATCCGAAGAAGAAATCTCATTAATCTGAA	659
OY	675	ataagtcgaagtgtttgagattgcactt 701	
Db	660	ATAAGTCAAGTGTTTGAGAATGCACCTT 686	

RESULT	7
LOCUS	AW952360
DEFINITION	AM952360 724 bp mRNA EST 01-JUN-2000
ACCESSION	EST364430 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence
VERSION	AW952360
KEYWORDS	AM952360.1 GI:8142042 EST.
SOURCE	human. Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 724) Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holtz, I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J. Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray Unpublished (2000)
TITLE	Contact: John Quackenbush The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 3528 Fax: 301 838 0208 Email: johnq@tigr.org Plate: 38 Seq primer: Reverse.
JOURNAL COMMENT	Location/Qualifiers 1..724 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="MAGE resequences, MAGB" /note="Vector: pBluescriptSm"
FEATURES	
source	
BASE COUNT	198 a 179 c 189 g 158 t
ORIGIN	


```

/clone_id="NH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site_1: Sfil (ggcgccctgcgc); Site_2: Sfil (ggccatcatggccc);
Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTCATG-3' and 3' adaptor
sequence: 5'-ATCTCAGGCGCCGAGCGCGCCGCATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)"

```

BASE COUNT	178 a	125 c	96 g	210 t
ORIGIN				

Query Match	18.8%;	Score 604;	DB 107;	Length 609;
Best Local Similarity	100.0%;	Pred. No. 5.6e-138;		
Matches 604; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	2264	agccatgccaagagagagccttgcacgcccccaagcctaagaaccagagaaaaatcaa	2323
Db	1	AGCCATGCCAGAGAGAGGCGCTTTGACCGGCCCCAGAGCTAAAAAGCACACAGAGAAAATCAA	60
QY	2324	atgcttcctactcagcgctgacccaacttctctagctgtgccagcgccccaccactcctgc	2383
Db	61	ATGCTTCTACTCAGCGCTGACCCCAACTTTTCTAGTGCCACGCGCCCAACCACTCCTGC	120
QY	2384	agtaaccacacacatcaccaactgcttctctccaaagtgatctgtatcttagttcat	2443
Db	121	AGTACCCACACCATCACCACTGCTTCTCTTCCAAGTGATCTGTATCTTAGTTCAAT	180
QY	2444	tatttcttltgatlgatatgacactatataaaatttcatlttgagaattctcaattgt	2503
Db	181	TATTTTCTTTGATTTGATGATGACACTATATAAAATTTTCATTGGAGAAATTCGCAATGTP	240
QY	2504	atctagttaaatagcacagtttggaaactgtctgagactgacttatacaataactaac	2563
Db	241	ATCTAGTTAAATAGCACAGTTTGGAAACTGTCTGAGACTGACTTTATCAATACTTAAC	300
QY	2564	cgacaaagatcatatccatgtgtatgtgttagacatttatttcatltgactaaaccag	2623
Db	301	CGACAAAGATCATATFCCATGTGTATGTGTTAGACATTTTATTTCATGTGACTTAACCCAG	360
QY	2624	gacagtttcagtgatgcaaatgtgtgcccctctggttcagctgaaacagtcctggaactt	2683
Db	361	GACAGTTTCAGTGATGCAAAATGTGTGCGCCCTCTGTGTTCAAGCTGAACACAGTCCGTGACTTT	420
QY	2684	caaaaaacctggaataagctcctccacacagttgtatataaattggacaatttaggaattttaaac	2743
Db	421	CAAAAACTTGAATAAGTCTCCACACAGTGTATATAAATTGGACAATTTAGGAATTTTAAAC	480
QY	2744	tttagatgatcatctgggttccatttctaattcattttaaatttltgttaatgcaaacagga	2803
Db	481	TTTAGATGATCATTTGGTTCATTTTATTTCAITTTTATTTTGTTAATGCAAAACAGGA	540
QY	2804	cttaaatgaacttgaatctctgtttaaagattatataaaaaacattgtgtatctatacat	2863
Db	541	CTTAATATGAACTTTGATCTCTGTTTAAAGATTATTTAAAAAACATTGTGTATCTATAACAT	600
QY	2864	atgg 2867	
Db	601	ATGG 604	

RESULT	11
BE613972	
LOCUS	BE613972 686 bp mRNA
DEFINITION	601504093F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5' /
ACCESSION	BE613972 mRNA sequence.

VERSION	BE613972.1	GI:9895569
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 686)	
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D.	

CDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: L1CM700 row: j column: 24
High quality sequence stop: 675.

FEATURES	Location/Qualifiers
source	1. .686

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3905783"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 2.1 kb."

```

BASE COUNT	226 a	156 c	196 g	108 t
ORIGIN				

Query Match	18.8%;	Score 603.8;	DB 110;	Length 686;
Best Local Similarity	97.1%;	Pred. No. 6.4e-138;		
Matches 669; Conservative	0;	Mismatches 12;	Indels 8;	Gaps 5;

OY	565	aacacgatgctgtcgcacaagcglttgagatccctgcagaatgagccccgtccagagagc	624
Db	1	AACACGATGCTGTGCACAAGCGTTGAGGATCCTGCAGATGAGCCCCCTGCCAGAGAGCC	60
OY	625	tggaggtgaattggaagagaaatccgaagaanaaatctcaataaactgaataagtcaa	684
Db	61	TGGAGGTGAATTGGAAGAGAATCCGAAGAANAATCTCAATAAATCTGAATAAGTCAAG	120
OY	685	tgtttgagattgcacttaaacggaacttgcctgtgaatttcgaggtgycgccgyagagt	744
Db	121	TGTTTGAGATTGCACCTTAACGGAACTTGCCCTGTAATTCGAGGTGGCCCGGAGAGTG	180
OY	745	gcccacccccacatgaaagaactttgtaccaaggtttcgttgygagtttgytgyggaag	804
Db	181	GCCCACCCCCACATGAAGAAC-TTGTGACCAGAAGTTTCGGT--GGGAGTTGTGGGGGAAG	237
OY	805	gtgaagggaaaaagcagaagaatttcaaaaaaatgcgcgcatagtctgttcttgagagc	864
Db	238	GTGAAGGGAAAAAGCAGAAGATTTCAAAAGAAAAATGCCGCCATAGCTGTCTTGAGAGAGC	297
OY	865	tgaagaagtlaccgccccctgcctgcaglttgaacgagtaaagcctaagaatcaaaaaagaaa	924
Db	298	TGAAGAAGTTACCGCCCCCTGCCTGCAGTTGAACGAGTAAGCCTAGAAATCAAAAAAGAAAA	357
OY	925	caaaaaccatagttcaagccacagacaagccagaatatgycaggygatcaatccgatla	984
Db	358	CAAAACCCATAGTCAAGCCAAGACACAGAACCCAGACGAATATGGCCA-GGGATCAATCCGATT	416
OY	985	gccgactggccagatccagcaagcaaaaaagagaagagccagagtlacaagctcctca	1044
Db	417	GCCGACTGGCCAGATCCAGCAGGCCAAAAAAGAGAGAGGCCAGATACCGCTCTCTCA	476

```
QY 1045 cagagcagagcctccgcgcgcgcagggagttgtgtatgcaggtgaaggttgaaaccaca 1104
      |||||||
Db 477 CAGAGCGAGGCCCTCCCGCGCAGGGAGTTGTGATGCAGGTGAAGTTGGAACACACA 536
      |||||||
QY 1105 ctgcagaaggaacggygcaccaacaagaaggtgycacaagcgcaatgcagccgagaacatgc 1164
      |||||||
Db 537 CTCGAGAAGGAACGGCGCACCAACAAGAAGGTGGCCAGCGCAATGCAGCCGAGAACATGC 596
      |||||||
QY 1165 tggagatccttgg-ttccaagtcgccgcgcgcgcgcgcaccaaccgcgcactcaagtca 1223
      |||||||
Db 597 TGGAGATCCTTGGTTTCAAGTCCCGCAGCGCGCAGCCACCAAA--CCGACTCAAGTCA 653
      |||||||
QY 1224 gagagaagacaccataaagaaccag 1252
      |||||||
Db 654 GAGGAGAAGACACCCATTAAGAAACAGGG 682
      |||||||

RESULT 12
AV707406 674 bp mRNA EST 09-OCT-2000
LOCUS AV707406 ADB Homo sapiens cDNA clone ADBCMG03 5', mRNA sequence.
DEFINITION AV707406
ACCESSION AV707406
VERSION AV707406.1 GI:10724671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 674)
AUTHORS Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA ADB clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source 1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADBCMG03"
/clone_lib="ADB"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 197 a 148 c 113 g 215 t 1 others
ORIGIN

Query Match 18.7%; Score 602.2; DB 38; Length 674;
Best Local Similarity 98.4%; Pred. No. 1.6e-137;
Matches 618; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
```

```
QY 2401 cactgtctctcttccaacagtgatcgtlatcttagttcaattatttcttttgatga 2460
      |||||||
Db 181 CACTGCTTCTCTTCCACACAGTGAATCTGATTTCTTAGTTCAATATTTCTTTGATTTGA 240
      |||||||
QY 2461 tatgacataataaaatttccatttgagaatttctcfaatgtatctagttaaataagcac 2520
      |||||||
Db 241 TATGACACTATATAAAATTTTCATTGGAGAAATTTCTCAATTGTATCTAGTTAAATAGCAC 300
      |||||||
QY 2521 agttggaacttgctcgtgactgacttatacaataatcaaccgacaagaatcatatcc 2580
      |||||||
Db 301 AGTTTGAAACTTGCTCTGAGACTGACTTTATCAATAATCTAACCGACAAGATCATATCC 360
      |||||||
QY 2581 atgtatgtggttagacatttatttcatctgactaaccagagacgtttcagtgatgc 2640
      |||||||
Db 361 ATGTGATGTGCTTAGACATTTTATTTTCATTGACTTAACCGACAGACAGTTTCAGTGTGC 420
      |||||||
QY 2641 aaattgtgcccctcgtgttcagctgaacacgttcctgacttcaaaaccttgaataag 2700
      |||||||
Db 421 AAATTGTGTGCCCTCTGTTTCAGCTGAACAGTCTCGACTTTCAAAAACCTTGAATAG 480
      |||||||
QY 2701 tctccacagttgtataaattgacaaatttagaattttaaacttagatgatcattgg 2760
      |||||||
Db 481 TCTCCACAGTTGTATAAATTTGACAAATTTAGGAATTTTAACTCTAGATGATCATTTGG 540
      |||||||
QY 2761 ttccattttatttcaatttatttatttgaatgaacacagacttaaatgaacttgat 2820
      |||||||
Db 541 NTCCA-TTTTATTTTCATTTCTTATTTTGGTAATGCAACAGAGCTTAATGAACTTTGAT 599
      |||||||
QY 2821 ctctgtttaagaattatataaaacat 2848
      |||||||
Db 600 CTCTGTTTAAAGAGTATTAAACAGAGAT 627
      |||||||

RESULT 13
BE674157 616 bp mRNA EST 08-SEP-2000
LOCUS BE674157
DEFINITION 7d76a04.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3278862 3'
similar to TR:O95793 O95793 STAU-FEN PROTEIN. ;, mRNA sequence.
ACCESSION BE674157
VERSION BE674157.1 GI:10034698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 616)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL, send email to:
info@image.lnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 469.

FEATURES
source 1..616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3278862"
/clone_lib="NCI-CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pRT73D-Pac (Pharmacia) with a
```

modified polylinker; plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 179 a 153 c 139 g 141 t 4 others

Query Match 18.5%; Score 594.4; DB 111; Length 616;
Best Local Similarity 98.4%; Pred. No. 1.3e-135;
Matches 598; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1230 aagacacccataaagaaacccaggagatggaagaaagtaacccttttgaacctgctc 1289
Db 8 AAGACACCCATAAAGAAACCAAGGAGTGAAGAAAGTAACCTTTTGAACCTGGCTCT 67
QY 1290 ggggatgaaatgggactagtaataaaggagatgagtcagatgccttataagtcat 1349
Db 68 GGGGATGAAAATGGGACTAGTAAATAAGAGATGAGTTCAGATGCCTTATCTAAGTCAT 127
QY 1350. cagcagctgcctgcggaattcttcccatggtcccgaggtcgcccgagctgtagagtt 1409
Db 128 CAGCAGCTGCTGCTGGAATTCTTCCCATGCTGCCGAGGTCGCCGAGGCTGTAGAGATT 187
QY 1410 agtcaaggacatcacaccaaagaatttaccaggcgagctccgaatcctctgccaaggccag 1469
Db 188 AGTCAAGGACATCACACCAAGATTTTACAGGGGACAGCTCCGAATCTGCCAAGGCCACG 247
QY 1470 gtaactgcatgatacccgagagatgttgatggyggcacctcgccacacagccgagacc 1529
Db 248 GTAAGTCCATGATAGCCCGAGAGTTGTTATGGGGGACCTCGCCACACAGCCGAGACC 307
QY 1530 attttaaagaatacatctctcagggcacgtaccatgacaccttcaacgagacctct 1589
Db 308 ATTTTAAAGATATACATCTCTTCCAGGGCCAGCTACCCCATGAGCTCTCACAGACCTCT 367
QY 1590 gagcaactgactatcttccagagtcacaggatccaggttgatatacaaaagacttccc 1649
Db 368 GAGCAACTGACTATCTTCCAGAGTCCAGGATTCACAGTTGAATCAAAAGACTTCCCC 427
QY 1650 aaaaacaagaagaattgtatatcttatacaattgctcctcagccaccttgatc 1709
Db 428 AAAAACAACAAGACGAATTTGTATCTCTATCAATTGCTCTCCANCCACCTGTATC 487
QY 1710 agccatggtatcggaagatgtgagtcctgcctcatgatatagtcgtcgtgaacatcta 1769
Db 488 AGCCATGGTATCAGCAAGATGTGAGTCTCTGCCATGATATGGCTGCGCTGAACATCTTA 547
QY 1770 aagttgctgtctgagttggaccaacaagaatcacagatgccaagaacaggaacgacca 1829
Db 548 GAGTTGCTGTCTGATGTGACCAACAANTACAGAGATGCCAAGACANGANACAGACCA 607
QY 1830 atgtctgt 1837
Db 608 ATGTCTGT 615

RESULT 14
AM162002/c 594 bp mRNA EST 09-NOV-1999
LOCUS au72c03.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION IMAGE:2781796 3' similar to TR:O95793 O95793 STAUFEN PROTEIN. ;,
mRNA sequence.
ACCESSION AM162002
VERSION AM162002.1 GI:6301035
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

1 (bases 1 to 594)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Other_ESTs: au72c03.y1
Contact: Wilson RK

TITLE

JOURNAL

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 464.

COMMENT

Location/Qualifiers
1. 594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2781796"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGAGCTCAAGGATCTTAATTAAATTATCCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGACTCGAGTTTCTTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."

FEATURES

source

BASE COUNT 130 a 152 c 123 g 188 t 1 others

ORIGIN

Query Match 18.3%; Score 588.2; DB 40; Length 594;
Best Local Similarity 99.3%; Pred. No. 4.3e-134;
Matches 590; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 324 gaaagcataccctactgtagaactaaatgcaactgcatgaaactggaaaaaacca 383
Db 594 GAAAGCATAAACCCCTACTGTAGACTANATGCACGTGCATGAACCTTGGAATAACCCA 535
QY 384 atgtataagcctgttgacccttactctcgatgcagtcaccactataactacaatgaga 443
Db 534 ATGTATAAGCCTGTGACCCCTTACTCTCGGATGCAGTCCACCTATTAACATGAGA 475
QY 444 ggaagtgctatccccgaggtacttcccaattccagttccaccttacttatcaa 503
Db 474 GGAGGTGCTTATCCCCCGAGGTACTTTTACCATTTCCAGTTCCACCTTTACTTTATCAA 415
QY 504 gtggaacttctgtgggagagacgaatlaatgycnaagaaagacaagacagctgcg 563
Db 414 GTGGAACCTTCTGTGGAGAGACAGCAATTAATGGCAAAAGAAAGACAAGACAGGCTGCG 355
QY 564 aaacacgatgtctgtccaaagcgttgagatcctgcagaatgagccctgccaagagag 623
Db 354 AAACACGATGCTGCTGCCAAAGCGTTGAGAGATCCTGCAGAAATGAGCCCTGCGCAGAGAG 295
QY 624 ctggaggtgaatgaaagaaatccgaaagaaagaatctcaataaatctgaaataagtaa 683

Db 294 CTGAGAGTGAATGGAAGAGATCCGAAGAGAAAATCTCAATAATCTGAATAAAGTCAA 235
QY 684 gtgtttgagattgcacttaaacggaacttgccctgtgaattcgaagttgcccggagagt 743
Db 234 GTGTTTGAGATTGCATTAAACGGAACCTTGCTGTAATTCGAGGTGGCCGGGAGAGT 175
QY 744 ggcacacccacatgaagaactttgtaccaagtttcggttgaggagttgtggggaa 803
Db 174 GGCACACCCACATGAAGAATTGTGACCAAGTTTCGGTTGGGAGTTGTGGGGAA 115
QY 804 ggtgaagggaagaagaagatttcaagaanaatgcgcacatagctgttcttgaggag 863
Db 114 GGTGAAGGGGAAAGCAAGAGATTTCAGAAGAAAATGCCGCATAGCTGTCTTGAGGAG 55
QY 864 ctgaagaagttaccgccccctgctgcagttgaaagtaaaagcctagaatcaaa 917
Db 54 CTGAAGAAGTTACC GCCCTGCTGCAGTTGAACGAGTAAGCCTAGAAAAAA 1

RESULT 15
LOCUS AL037151 669 bp mRNA EST 29-FEB-2000
DEFINITION DKFZp564O0264_r1 564 (synonym: hfbr2) Homo sapiens cDNA clone
DKFZp564O0264 5', mRNA sequence.
ACCESSION AL037151
VERSION AL037151.3 GI:5928071
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 669)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
S.
TITLE EST (Duesterhoeft, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5866472.
Contact: Duesterhoeft A

MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp564O0264) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1..669
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp564O0264"
/clone_lib="564 (synonym: hfbr2)"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="xl-2blue"
/note="Vector: PAMPl; Site_1: NotI; Site_2: SalI"
BASE COUNT 163 a 212 c 143 g 147 t 4 others
ORIGIN

Query Match 17.9%; Score 576; DB 28; Length 669;
Best Local Similarity 95.2%; Pred. No. 4.5e-131;
Matches 636; Conservative 0; Mismatches 24; Indels 8; Gaps 4;

QY 11 gggctgagggcgagcgtgagc-gtcttcagcgtttgagc--ggcggtgagcgtctctc 67
Db 2 GGGCTGGGGGGCGCTGAGCGGCTCTTCAAGCTTGGCGCGGGGCTGCCGCGTCTCTC 61
QY 68 ggtcccgcttccttgaccgctcccccccgccggcgcgcgccgctctctcaag 127

Db 62 GGTCCCGCTTCCTTTGACCCGCTCCCCCCCCCGGGCGGGCGCCGCTCTCCACG 121
QY 128 gccactcgccctctccctccctcgctccctctctctccctctccctctctccct 187
Db 122 GCCACTCCGCGCTTCCTCCCTCCCTGCTCCCTTCCTCTCCCTTTTTCCTCTCCTT 181
QY 188 cccctctcgccgcaaccgcccagaccgcccggggagagctcgagcaagcacc 247
Db 182 CCCCTCTCGCCGCCACCGCCGAGACCGCGCGGGGAGCAGCTCGGACGACGACC 241
QY 248 agagttatataaccacttaacctctcagaactgaaacaaagacaacattgttctctggaag 307
Db 242 AGAGTTATTAACCACTTAACCTCTCAGAACTGAACAAGACAACATTTGTTCTGGAACG 301
QY 308 cccctctttaaaaaaagaagataaaccctactgtgaaactaaatgcactgtgcatgaa 367
Db 302 CCCTCTTTTAAAAAAGAAAGCATTAACCCCTACTGTAGAACTAAATGCACCTGTGCATGAA 361
QY 368 acttgaaaaaaacccaatgtataagcctgttgaccctactctcgaatgcagtcaccta 427
Db 362 ACTTGAAAAAAACCAATGTATAAGCCTGTTGACCCCTTACTCTCGGATGCAGTCAACCAA 421
QY 428 taactacaacatgagagagtgcttatcccccgaggtactttaccatttcagttcc 487
Db 422 TAACCTACACATGAGAGAGGTGCTTATCCCCCGAGGTACTTTTACCATTTCAGTTC 481
QY 488 acccttaacttatcaagtgaactttctgtgaggagacagcaatttaatgcaaaaggaa 547
Db 482 ACCTTTACTTTATCAAGTGAAGTCTGTGGAGGACAGCAATTTAATGGCAAGGAAA 541
QY 548 gacaagacagctgcgaaacacagatgctgtgccaaagcgttgagatcctgcagaatga 607
Db 542 GACAAGACAGGCTGCGAAACACGATGCTGCTGCCAAAGCGTTGAGGATCCTGCAAAATGA 601
QY 608 gccctgcccagagaggtgg--agtgaaatggaagagaatcc--gaagaagaatctc 662
Db 602 ANCCCTGCCANAAAGGCTGGGAAGTGAATGGAANAGAAATCCCGAAGAAANAAATCTC 661
QY 663 aataaatc 670
Db 662 AATAAATC 669

Search completed: April 5, 2001, 00:29:42
Job time: 53078 sec

US-08-458-568A-11/C
; Sequence 11, Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; APPLICANT: Yeh, Lily
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
; US-08-458-568A-11

Query Match 1.4%; Score 44.2; DB 1; Length 12001;
Best Local Similarity 58.9%; Pred. No. 0.086;
Matches 76; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 90 ctcccccccgccgagcgagcgagcctctcaccagccacatccgctcttccctcc 149
Db 7697 CCCCCCTCCTCGCCCGCGGTCCCTCCGCGCGCGGTCCCTCCGCGCGCG 7638
QY 150 ttctcct 209
Db 7637 CGCGTCCCGCTCTCTCGCGCGCGCGGTCCCTCCGCGCGCGGTCCCTCT 7578
QY 210 aggaccgac 218
Db 7577 CCGCCACAC 7569

RESULT 12
US-08-323-443B-1
; Sequence 1, Application US/08323443B
; Patent No. 5654170
; GENERAL INFORMATION:

APPLICANT: KLINGER, KATHERINE W.
; APPLICANT: LANDES, GREGORY M.
; APPLICANT: BURN, TIMOTHY C.
; APPLICANT: CONNORS, TIMOTHY D.
; APPLICANT: DACKOWSKI, WILLIAM R.
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,443B
; FILING DATE: 12-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0372/0A462
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: PKD1 GENOMIC
; US-08-323-443B-1

Query Match 1.3%; Score 41.4; DB 1; Length 31571;
Best Local Similarity 53.1%; Pred. No. 0.9;
Matches 111; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 1 actctcgc 60
Db 9941 ACTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9998
QY 61 ctctctcgc 120
Db 9999 CCTCCACACCT 10058
QY 121 ctcaagc 180
Db 10059 CACCT 10118
QY 181 ctctctcgc 209
Db 10119 TCCCTCCCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10147

RESULT 13
US-08-781-891-79
; Sequence 79, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:


```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
5  SOFTWARE: Wordperfect (Version 5.1)
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/452, 872
8  FILING DATE:
9  CLASSIFICATION:
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US/08/147, 777
12 FILING DATE:
13 APPLICATION NUMBER: 07/876, 289
14 FILING DATE: April 30, 1992
15 APPLICATION NUMBER: Unassigned (204/144)
16 FILING DATE: October 29, 1993
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Wardburg, Richard J.
19 REGISTRATION NUMBER: 32,327
20 REFERENCE/DOCKET NUMBER: 204/153
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (213) 489-1600
23 TELEFAX: (213) 955-0440
24 TELEX: 67-3510
25 INFORMATION FOR SEQ ID NO: 3:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 24979 base pairs
28 TYPE: nucleic acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 MOLECULE TYPE: DNA (genomic)
32 US-08-452-872-3

```

Query Match	1.3%	Score 40.8;	DB 3;	Length 24979;
Best Local Similarity	59.7%;	Pred. No. 1.1;		
Matches 86; Conservative	0;	Mismatches 57;	Indels 1;	Gaps 1.;

QY	54	tgcgcgctctctctcgcgtcccgcttccctltgacacgcctccccccccgcgcgcgcgc	113
Db	6957	TGCTCCTCTGGCTCTCTCTCCGGGCTCCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT	7016
QY	114	ccgcctctctcc-acggcgaactcgcctcttccctccctctgccttcttccctccct	172
Db	7017	GCTCCT	7076
QY	173	tttctctcttcttcttccctctctc	196
Db	7077	TTCTCCTCTTTTCTTCTTCTTCTC	7100

Search completed: April 5, 2001, 08:56:16
Job time: 58650 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 09:30:11 ; Search time 313.08 Seconds
(without alignments)
1470.662 Million cell updates/sec

Title: US-09-316-048-9
Perfect score: 2857
Sequence: 1 cgggcggcgccctcccccgcg.....ccaataatgtaataactaa 2857

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.2	2.2	7218	1	US-08-232-463-14 Sequence 14, Appl
2	64	2.2	7218	1	US-08-232-463-14 Sequence 14, Appl
3	49.6	1.7	53526	3	US-08-658-136-2 Sequence 2, Appl
4	49.6	1.7	53577	3	US-08-658-136-1 Sequence 1, Appl
5	45.8	1.6	3918	3	US-08-836-329-1 Sequence 1, Appl
6	43.6	1.5	1946	2	US-08-861-464-1 Sequence 1, Appl
7	43.6	1.5	1946	2	US-08-396-001-1 Sequence 1, Appl
8	43.4	1.5	12001	1	US-08-458-568A-11 Sequence 11, Appl
9	42.2	1.5	1931	2	US-09-130-114-2 Sequence 2, Appl
10	41.8	1.5	1308	3	US-09-188-930-248 Sequence 248, App
11	41.8	1.5	1421	3	US-09-188-930-244 Sequence 244, App
12	40.6	1.4	2580	3	US-09-050-863-2 Sequence 2, Appl
13	40.6	1.4	5452	2	US-09-130-114-1 Sequence 1, Appl
14	40.6	1.4	10596	1	US-07-884-811-15 Sequence 15, Appl
15	40.6	1.4	10596	1	US-07-885-971-15 Sequence 15, Appl
16	40.6	1.4	10596	1	US-08-087-783A-15 Sequence 15, Appl
17	40.6	1.4	10596	1	US-08-087-783A-15 Sequence 15, Appl
18	40.6	1.4	10596	2	US-08-194-087-15 Sequence 15, Appl
19	40.6	1.4	10596	4	PCT-US93-04648-15 Sequence 15, Appl
20	40.2	1.4	1796	1	US-08-417-103-11 Sequence 11, Appl
21	39.2	1.4	987	1	US-08-230-047-6 Sequence 6, Appl
22	38.6	1.4	1796	1	US-07-816-283-11 Sequence 11, Appl
23	37.6	1.3	803	1	US-07-928-611-12 Sequence 12, Appl
24	37.6	1.3	803	2	US-08-487-811A-12 Sequence 12, Appl
25	37.6	1.3	803	4	PCT-US93-07370-12 Sequence 12, Appl
26	37.6	1.3	1610	1	US-08-056-051-5 Sequence 5, Appl
27	37.6	1.3	1610	1	US-07-928-611-21 Sequence 21, Appl
28	37.6	1.3	1610	2	US-08-487-811A-21 Sequence 21, Appl

29	37.6	1.3	1610	4	PCT-US93-07370-21	Sequence 21, Appl
30	37.2	1.3	8438	1	US-07-945-283-1	Sequence 1, Appl
31	37	1.3	421	2	US-08-332-766A-25	Sequence 25, Appl
32	37	1.3	9551	1	US-08-056-200-93	Sequence 93, Appl
33	37	1.3	9551	2	US-08-800-644-93	Sequence 93, Appl
34	36.8	1.3	24979	2	US-08-147-777-3	Sequence 3, Appl
35	36.8	1.3	24979	3	US-08-452-872-3	Sequence 3, Appl
36	36.8	1.3	24979	4	PCT-US93-03985-3	Sequence 3, Appl
37	36.2	1.3	12412	1	US-08-390-878-18	Sequence 18, Appl
38	36.2	1.3	152331	3	US-09-128-155-16	Sequence 16, Appl
39	36	1.3	659	1	US-07-928-611-10	Sequence 10, Appl
40	36	1.3	659	2	US-08-487-811A-10	Sequence 10, Appl
41	36	1.3	659	4	PCT-US93-07370-10	Sequence 10, Appl
42	36	1.3	1466	1	US-08-056-051-3	Sequence 3, Appl
43	36	1.3	1466	1	US-07-928-611-19	Sequence 19, Appl
44	36	1.3	1466	2	US-08-487-811A-19	Sequence 19, Appl
45	36	1.3	1466	4	PCT-US93-07370-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELEPHONE: (703)836-9300
TELECOMMUNICATION INFORMATION:
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEO ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-fls
US-08-232-463-14
Query Match 2.2% Score 64.2; DB 1; Length 7218;

Best Local Similarity 2.5%; Pred. No. 6.4e-09;
Matches 6; Conservative 166; Mismatches 69; Indels 0; Gaps 0;

QY 10 cccctccccgtcactctctcgtccagctgagcccgagccgctctcagcgttgcgc 69
Db 1206 YY 1265
QY 70 tggctgtcgtcgtctgtgtgctgcccccttctctgagccccgctgagcgccc 129
Db 1266 YY 1325
QY 130 gcttcgcctccgcaactcgcctctcctcctctgtcgcgcccttctcctgcgcgt 189
Db 1326 YY 1385
QY 190 cttcactgtctctcactcctcctcgcgcgcccccaagaccgccccggagagctc 249
Db 1386 YYGTACCAAAATTC 1445
QY 250 t 250
Db 1446 T 1446

RESULT 2

US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: PT29pt-F1s

US-08-232-463-14

Query Match

2.2%; Score 64; DB 1; Length 7218;
Best Local Similarity 3.4%; Pred. No. 7.4e-09;
Matches 13; Conservative 227; Mismatches 142; Indels 0; Gaps 0;

QY 458 gggcgagcagcagtttaatggaagaagaatgagaccaccgctgaacacgatgcc 517
Db 1421 RRR 1362
QY 518 tggcgtgctgagagactctgcagagtgaaccctgcagaaaggttgagtaattgc 577
Db 1361 RRR 1302
QY 578 aagaagcagaggaagaagaacatcaataaalcgaataaagcgaagtgltgaattgc 637
Db 1301 RRR 1242
QY 638 gctgaagcgaatttgctgtgaatttgaggtgccccggagagtgccaccacacat 697
Db 1241 RRR 1182
QY 698 gaagaacttgtagccagaggttccagtggggaattgtgaggggaaggagaagaaag 757
Db 1181 RRR 1122
QY 758 caagaagatcacaagaagaatgcggccagggctgtctgagcagccttagaggtgcc 817
Db 1121 RRR 1062
QY 818 acccctccctgtgtgagcga 839
Db 1061 AAGCTCCTCGACCTGCAGCCA 1040

RESULT 3

US-08-658-136-2
Sequence 2, Application US/08658136
Patent No. 6071717

GENERAL INFORMATION:

APPLICANT: KLINGER, KATHERINE W

APPLICANT: LANDES, GREGORY M

APPLICANT: BURN, TIMOTHY C

APPLICANT: CONNORS, TIMOTHY D

APPLICANT: DACKOWSKI, WILLIAM

APPLICANT: GERMINO, GREGORY

APPLICANT: QIAN, FENG

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENZYME CORPORATION

STREET: ONE MOUNTAIN ROAD

CITY: FRAMINGHAM

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 01701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/658,136

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: LASSEN, ELIZABETH

REGISTRATION NUMBER: 31,845

REFERENCE/DOCKET NUMBER: GEN4-17.8

TELECOMMUNICATION INFORMATION:

TELEPHONE: 508-872-8400

TELEFAX: 508-872-5415

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 508 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: PT29pt-F1s

US-08-658-136-2


```
Db 101 CCGCCAGAGCGCGCCAGAGCCGCTCTCCAGCCGCGCGGCTCGCCACCTAGACCCC 160
QY 71 ggcgtgctgcgtctgtgtgctcctcctctctgagcccgctgagcgcccg 130
Db 161 CCCAGCCACCCCTTCTCCGCGCGCGCGCGCTCTCTCCCGCGCGCGCGCG 220
QY 131 ccttcgctcgcgaactcgcctctcctcctctgtgagcccttttctcgcgtc 190
Db 221 CCCCCCTCTCTCCCGCGCGCGCTCGCTCCCGCTCTCTCTCTCTCCACACCGCC 280
QY 191 ttcactgtctctcactcctcctcgcgcgcgcgaaga 227
Db 281 CTCAGCCGCTCTCTCTCTAGCGCGCTCTGAAGAAGA 317
```

RESULT 6

```
US-08-861-464-1
; Sequence 1, Application US/08861464
; Patent No. 5874210
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/861,464
; FILING DATE: 22-MAY-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/396,001
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09351
; FILING DATE: 15-AUG-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,408
; FILING DATE: 16-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 322..1671
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 322..1671
US-08-861-464-1
```

Query Match 1.5%; Score 43.6; DB 2; Length 1946;
Best Local Similarity 54.3%; Pred. No. 0.005;
Matches 88; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
QY 42 gccccgagccgctcttcagcglttgagctgctgctgctgtgtgtgctccct 101
Db 733 GCCACACCGCTACTCGCGGCTGTGATGAAGATGAGTGTGTGTGCTGCGCAC 792
QY 102 tctctgagcccgctgagcgccgcttcgctccgcaactccgctctcctc 161
Db 793 ACTACTTGTCTCTGACTGCGCGCGCGCTACTACTCTGCGCGCGCTCTCTCC 852
QY 162 ctctgctgagcccttttctcctgagctctcactgtctct 203
Db 853 TCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 894
```

RESULT 7

```
US-08-396-001-1
; Sequence 1, Application US/08396001
; Patent No. 5919618
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence In
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 322..1671
US-08-396-001-1
```

Query Match 1.5%; Score 43.6; DB 2; Length 1946;
Best Local Similarity 54.3%; Pred. No. 0.005;
Matches 88; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
QY 42 gccccgagccgctcttcagcglttgagctgctgctgctgtgtgtgctccct 101
```

[illegible]

RESULT 8
US-08-458-568A-11
; Sequence 11, Application US/08458568A

```

: GENERAL INFORMATION:
: APPLICANT: Schaffer, Priscilla A.
: APPLICANT: Yeh, Lily
: TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
: TITLE OF INVENTION: Infections
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,568A
: FILING DATE: 02-JUNE-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/065,146
: FILING DATE: 05-MAY-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary Ph.D., Kathryn R.
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: DFCI-0029
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12001 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Herpes simplex virus
: STRAIN: Herpes Simplex Virus Type 1
:
: US-08-458-568A-11

```

Query Match	1 5%;	Score 43.4;	DB 1;	Length 12001;
Best Local Similarity	47.9%;	Pred. No. 0.02;		
Matches 125; Conservative	0;	Mismatches 136;	Indels 0;	Gaps 0;

QY 2 ggagcgcgccacctccccgcacttcctcgtccacgctgcgggcccccagaccgctcttcagc 61
 ||| | ||||| - | | | | | | | | | | | | | | | |
Db 1210 gggggggggcccgcttttggcgcttgccgcgctccctcccccgcctcccccggctctcc 1269

QY 62 gttgcgctgctgtcgtcgctctgtgtgcgcctcaccctctctctgagcccgctg 121

[illegible]

QY 242 acgaactctggygaacagcc 262
| | | | | | | | | |
Db 1450 ccccgctccgcggcccccgc 1470

```

RESULT      9
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D9030US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

Query Match	1.5%;	Score 42.2;	DB 2;	Length 1931;
Best Local Similarity	48.5%;	Pred. No. 0.013;		
Matches 116; Conservative	0;	Mismatches 123;	Indels 0;	Gaps 0;

[illegible]

RESULT 10
US-09-188-930-248/c
; Sequence 248, Application US/09188930A

```

; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murlson, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A

```

```

; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 248
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-248

```

Query Match	1.5%;	Score 41.8;	DB 3;	Length 1308;
Best Local Similarity	51.5%;	Pred. No. 0.013;		
Matches 120; Conservative	0;	Mismatches 112;	Indels 1;	Gaps 1;

[illegible]

RESULT 11
US-09-188-930-244/c
; Sequence 244, Application US/09188930A

```

: APPLICANT: Watson, James D.
: APPLICANT: Strachan, Lorna
: APPLICANT: Sleeman, Matthew
: APPLICANT: Onrust, Rene
: APPLICANT: Murison, James Greg
: TITLE OF INVENTION: Compositions Isolated From Skin Cells
: TITLE OF INVENTION: and Methods For Their Use
: FILE REFERENCE: 11000.1011c1
: CURRENT APPLICATION NUMBER: US/09/188,930A
: CURRENT FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 348
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 244
: LENGTH: 1421
: TYPE: DNA
: ORGANISM: Mouse
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1370)...(1370)
: NAME/KEY: unsure
: LOCATION: (1395)...(1395)
US-09-188-930-244

```

Query Match	1.5%;	Score 41.8;	DB 3;	Length 1421;
Best Local Similarity	51.5%;	Pred. No. 0.014;		
Matches 120; Conservative	0;	Mismatches 112;	Indels 1;	Gaps 1;

QY	38	gcgagcccccgaagccgcctcttcagcgtttgcgcgtgcgtgcgtctgtgtgcgtcc	97
Db	893	GCTGGCTGCGGGAAACCCCTTCAGCCTCTGTGATGAGGCTGTCTCTCAGGGACCTCCGGCTT	834
QY	98	ccctctcttgaagccccgcgcgtgcgcgcgcc-gccttcgccttcgccactccgcctctt	156
Db	833	CTCTTCAGTGAACCCACACATACTGCGGCTTCCCGGGGCTCTCTCTGCGCCACTGAACACCTTC	774
QY	157	ccctccctctgtctgcgcgcccttttctctgcgcgcgtcttcacttgccttattcaactctgcgc	216

Db 773 CTCTTCCTTCTCTCCTCCTCCTCCTCCGAGCACCCCTTCACTGTATTGTCCCTCCCCACT 714

RESULT 12
US-09-050-

US-09-050-863-2/c
; Sequence 2, Application US/09050863

```

: GENERAL INFORMATION:
: APPLICANT: Lao, Ying
: APPLICANT: Hiang, Betty
: APPLICANT: Payan, Don
: TITLE OF INVENTION: Mammalian Protein Interaction Cloning
: TITLE OF INVENTION: System
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/050,863
: FILING DATE: 30-MAR-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REGISTRATION NUMBER: 38,304
: REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 949-8711
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2580 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
:
US-09-050-863-2

```

Query Match	1.4%;	Score 40.6;	DB 3;	Length 2580;
Best Local Similarity	54.1%;	Pred. No. 0.05;		
Matches 126; Conservative	0;	Mismatches 104;	Indels 3;	Gaps 2;

QY	6	g c g g c c c t c e c c c g l t a c t t c o b t y c c a g c t g c g g g c c c c a g a c c g c t c t t a a g c g t t t c	65
Db	961	G C C C C T C T G C C C C T C C T C T G C T C T G C C C C T C C T G C C C C T C C T G C T C T G C C C C T	902
QY	66	g c g c t g c t g t c g t c g c g t c t g t g t g c g t c c c c c t c t t c t g a g c c c c g g c g c t g c g a c	125
Db	901	C C T C C T G C T C C T G C C C C C T C C T G C C C C T C C T G C C C C T C C T G C C C C T C C T G C	843
QY	126	g c c c g c c t t g c c t c c g c c a c t c c g c c t c t t c c c t c c t c t g t f c g g c c c c t t t t c c t g	185
Db	842	C C C T C C T C C T G C C C C C T C C T G C C C C C T C C T G C C C C T C C T G C C C C T C C T G C C C C T	785
QY	186	c c g t c t c a c t g c t t c t t c a c c t c o c t c g c c g c c c a g a c c g c c g c c c c c	238
Db	784	C C T C C T G C T C C T G C C C C C T C C T G C T C C T G C C C C C T C C T G C C C C C T C C T G C C C C C	732

RESULT 13
US-09-130-114-1
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/ID903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
US-09-130-114-1

Query Match 1.4%; Score 40.6; DB 2; Length 5452;
Best Local Similarity 54.1%; Pred. No. 0.084;
Matches 126; Conservative 0; Mismatches 104; Indels 3; Gaps 2;

QY 6 gcggcccccctcccccgtcaactctctccagcgtcgagggcccccagccgcctctcagcgtt 65
Db 1844 gccctctctgcccctctctctctctctctgcccctctctctctctctctctctgcccct 1903
QY 66 gcgctgctgtcgtcggtctgtgtgcgtcccccctctctctagccccggcctgagcgc 125
Db 1904 cctctctgctctgcccctctctctgcccctctctctctctctctctctctctctctgc 1962
QY 126 gccgcctctgcctccgcacatccgcctctctctctctctctctctctctctctctctg 185
Db 1963 cctctctctgctctctgcccctctctctctctctctctctctctctctctctctctgcccct 2020
QY 186 ccgtcttcaactgtctcttcaactctctctgcccgcgcgcccagaccgcgcgcgc 238
Db 2021 cctctctgctctgcccctgcccct 2073

RESULT 14
US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-884-811-15

Query Match 1.4%; Score 40.6; DB 1; Length 10596;
Best Local Similarity 54.1%; Pred. No. 0.13;
Matches 126; Conservative 0; Mismatches 104; Indels 3; Gaps 2;

QY 6 gcggcccccctcccccgtcaactctctccagcgtcgagggcccccagccgcctctcagcgtt 65
Db 2541 GCCCTTCCTGCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 2482
QY 66 gcgctgctgtcgtcggtctgtgtgcgtcccccctctctctagccccggcctgagcgc 125
Db 2481 CCTCCTGCTCCTGCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 2423
QY 126 gccgcctctgcctccgcacatccgcctctctctctctctctctctctctctctctctg 185
Db 2422 CCT 2365
QY 186 ccgtcttcaactgtctcttcaactctctctgcccgcgcgcccagaccgcgcgcgc 238
Db 2364 CCTCCTGCTCCTGCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 2312

RESULT 15
US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases

OY 1741 tgcacaagaacaggaatatgacacagtttcacgctgcggaggtgtcgaacctttctgccc 1800
|||||
Db 1742 tgcacaagaacaggaatatgacacagtttcacgctgcggaggtgtcgaacctttctgccc 1801
OY 1801 acaaacattataaaacccaacatatatactagtaaaatactgagaactgctttgaaattc 1860
|||||
Db 1802 acaaacattataaaacccaacatatatactagtaaaatactgagaactgctttgaaattc 1861
OY 1861 ggaatatctgataaactccagtgggccaaagacatggtgataaaaatgtggcaaaagacgac 1920
|||||
Db 1862 ggaatatctgataaactccagtgggccaaagacatggtgataaaaatgtggcaaaagacgac 1921
OY 1921 aagaaactcaggtgtagccctggtgtgctgycggctagtgatgactgtgtctgc 1980
|||||
Db 1922 aagaaactcaggtgtagccctggtgtgctgycggctagtgatgactgtgtctgc 1981
OY 1981 catccatccagacagaaacacagccccaagcctccagttctgttttgcacgcgtgacaaa 2040
|||||
Db 1982 catccatccagacagaaacacagccccaagcctccagttctgttttgcacgcgtgacaaa 2041
OY 2041 gagagcacagccaattctcatgtgcttcttcagatactttgaaaaacccgacagcca 2100
|||||
Db 2042 gagagcacagccaattctcatgtgcttcttcagatactttgaaaaacccgacagcca 2101
OY 2101 caccagagagccttatagcggcccggaactaaacgagaccagagaaaaaggccagtgctt 2160
|||||
Db 2102 caccagagagccttatagcggcccggaactaaacgagaccagagaaaaaggccagtgctt 2161
OY 2161 cctactgcacatgaactgaactcagctccgcacacagtagcaccactgtaaccaactgtcttc 2220
|||||
Db 2162 cctactgcacatgaactgaactcagctccgcacacagtagcaccactgtaaccaactgtcttc 2221
OY 2221 tcttc-agtttcatttttttcccttgatgatacaacaactataatttcatttcagttcc 2279
|||||
Db 2222 tcttcagtttcatttttttcccttgatgatacaacaactataatttcatttcagttcc 2281
OY 2280 ttagtcgtgtctacttaactagcagtttagaactgtcagtcagtgaactgccaagatc 2339
|||||
Db 2282 ttagtcgtgtctacttaactagcagtttagaactgtcagtcagtgaactgccaagatc 2341
OY 2340 acagcccggttggtggtgcatctctgtgcctcctggtctgtgtaacagttctggaattacca 2399
|||||
Db 2342 acagcccggttggtggtgcatctctgtgcctcctggtctgtgtaacagttctggaattacca 2401
OY 2400 ccagaatccttgactccctgcgcctgtgataaattggacagccttagaacttttaactt 2459
|||||
Db 2402 ccagaatccttgactccctgcgcctgtgataaattggacagccttagaacttttaactt 2461
OY 2460 agatcaaaaagatatgtccttttaactttaatttaagagcagactttaaatgagcc 2519
|||||
Db 2462 agatcaaaaagatatgtccttttaactttaatttaagagcagactttaaatgagcc 2521
OY 2520 ctgaccttaccattataacagaattgtgcataaaaggagtgtttcttgaggagtagctt 2579
|||||
Db 2522 ctgaccttaccattataacagaattgtgcataaaaggagtgtttcttgaggagtagctt 2581
OY 2580 tttttaccacactacagacattactgttagggccagaagactacaggtgtgtccct 2639
|||||
Db 2582 tttttaccacactacagacattactgttagggccagaagactacaggtgtgtccct 2641
OY 2640 agagggcccaataacagtcactccaactctaagtcggggaaggtgacaggttccctgg 2699
|||||
Db 2642 agagggcccaataacagtcactccaactctaagtcggggaaggtgacaggttccctgg 2701
OY 2700 tgcctggtgtgcacaggggcagggcaggtcagctggcctgagggaagagcatgtgctcccta 2759
|||||
Db 2702 tgcctggtgtgcacaggggcagggcaggtcagctggcctgagggaagagcatgtgctcccta 2761
OY 2760 gtgcagccctgtctcactcttggtttagctggaaccttcccaactcagtgaataaagta 2819
|||||
Db 2762 gtgcagccctgtctcactcttggtttagctggaaccttcccaactcagtgaataaagta 2821

OY 2820 aactcaacttcttgttcaccaataaatggtactactaa 2857
|||||
Db 2822 aactcaacttcttgttcaccaataaatggtactactaa 2859

RESULT 2
ID 293324 standard; cDNA; 3142 BP.
XX
AC 293324;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human staufen cDNA.
XX
KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 288..1193
FT /*tag= a
FT /product= Staufen protein
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83108.
XX
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
XX
PS Claim 4; Figure 1b; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hSP (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
SQ Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 47.4%; Score 1354.6; DB 21; Length 3142;
Best Local Similarity 84.6%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 269; Indels 21; Gaps 6;

[illegible]

Db	1078	ccgagaacatgctgtagatctcctggtttccaagtcccgacgycgcagccaccacaacccg	1137
QY	1147	catccaatcagaagagaagatccagttaaagaaacccagagacggaagaaagttaacgt	1206
Db	1138	caatcaagtcagagagaagacaccataaagaacccagggatgtaagaaaagtaacct	1197
QY	1207	tttttgaacctagccctggygatgaaaaatggaactagtaacaaagacgagaggttcaaga	1266
Db	1198	tttttgaacctgctctggygatgaaaaatggaactagtaataaagatgattcaaga	1257
QY	1267	tgccttatcttagccatcagcagctgcagctggaattctcccatggtgcccgaagt	1326
Db	1258	tgccttatctaagtcatacagcagctgcctgctggaattctcccatggtgcccgaagt	1317
QY	1327	cccaagctgtcggygttagtccaagacacacacacaaagatttaccagggcagctccaa	1386
Db	1318	cccaagctgtagaggttagtccaagacatcacaccaaagatttaccagggcagctccga	1377
QY	1387	atcctgcceaagcgaacggttaactgcatgatagcccgaagttgtgtacggygcacct	1446
Db	1378	atcctgcceaagcgcacggttaactgcatgatagcccgaagttgtgtatggygcacct	1437
QY	1447	cgcccacagccgagagaccattttaaagatlaacatctcttcagggccacgtaccccaatg	1506
Db	1438	cgcccacagccgagagaccattttaaagatlaacatctcttcagggccacgtaccccaatg	1497
QY	1507	ctgcgcaactagaccctctgagcaactgttactaccttccaagagcccagggatccaagtt	1566
Db	1498	ctcttcacgagaccctctgagcaactgtaactcttccaagagttcagggatccaagtt	1557
QY	1567	aatacaaaagatttcccaagaacaacaagaacgagttgtatctctcatcaactctcct	1626
Db	1558	aatacaaaagatttcccaagaacaacaagaacgaaattgtatctctcatcaactctcct	1617
QY	1627	cacagccgctctcgtcagtcagtcagtcgcgcaagagatgtagtccgtgcatagtatg	1686
Db	1618	ctcagccacctctgatacgcagcatgtatcgcgaagatgtgagttcctgcatagtatg	1677
QY	1687	ctgcactgaacattttaagctgtgtctgagtgtgacaacagagcacagagatgcca	1746
Db	1678	ctgcgctgaacatctttaagttgtgtctgagtgtgacaacaaagtaacagagatgcca	1737
QY	1747	gaacaggaatggaaccagtttcagcgtgcggygaggtgctgaaccttctgcccacaac	1806
Db	1738	gaacaggaagaacggaaccaatgtctgtgtggygaggtgctgaaccttctgcccataac	1797
QY	1807	caattataaaa-cccaacatatatactgaaaataactgagaactgctttgaaaatttgaat	1865
Db	1798	cattataaaaatcccaacatatatactgaaaataactga-aaactgctttgaaaatttgaat	1856
QY	1866	atctgataactccagtggygcgaagaca 1892	
Db	1857	ttctgataacctccagtggygcgaagaga 1883	

RESULT	3	
Z93323		
ID	Z93323	standard; cDNA; 3217 BP.
XX		
AC	Z93323;	
XX		
XX		
DT	04-JUL-2000	(first entry)
XX		
DE		
XX	Human stauflen cDNA.	
XX		
KW	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW	acquired immune deficiency syndrome; protease; human; ss.	
XX		
OS	Homo sapiens.	
XX		
XX		
PH	key	Location/Qualifiers

FT	CDS	363..1853
FT	/**tag= a	
FT	/product= Staufen	
XX		
PN	CA2238656-A1.	
XX		
PD	22-NOV-1999.	
XX		
Pf	22-MAY-1998; 98CA-2238656.	
XX		
PR	22-MAY-1998; 98CA-2238656.	
XX		
PA	(UYMO-) UNIV MONTREAL.	
PI	Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;	
PI	Mouland AJ;	
XX		
DR	WPI; 2000-246924/22.	
DR	P-PsDB; Y83108.	
XX		
PT	Isolated mammalian or Caenorhabditis elegans staufen proteins useful	
PT	for treating retroviral infections especially human immunodeficiency	
PT	virus infections (HIV)	
XX		
PS	Claim 4; Figure 1a; 96pp; English.	
XX		
CC	Staufen is a RNA binding protein which interacts with double stranded	
CC	RNA and/or rough endoplasmic reticulum. It is involved in mRNA	
CC	transport and localization. Mammalian SP contain multiple ds	
CC	RNA-binding domains. Chimeric proteins comprising portions of	
CC	staufen with proteins with RNase or protease activity can be used to	
CC	treat virus infections. The RNase or protease activity of the fusion	
CC	protein prevents proper maturation of the virus. Mammalian staufen	
CC	proteins seem to recognise double stranded RNA structure rather than	
CC	any sequence specific position. 2 bacterially-expressed fusion	
CC	proteins used in an RNA-binding assay (his/hsp (human SP) and	
CC	MBP/msp (murine sp)) both proteins strongly bound double stranded	
CC	RNA. Both fusion proteins also directly bound labeled ds RNAs and	
CC	RNA/DNA hybrids. Staufen proteins may therefore be useful in the	
CC	treatment of RNA virus infections, especially those caused by	
CC	retroviruses, in particular human immunodeficiency virus. This	
CC	sequence is an alternatively spliced sequence to the one given in	
CC	Z93322.	
XX		
SQ	Sequence 3217BP; 889 A; 796 C; 742 G; 790 T; 0 other;	

Query Match	44.48;	Score 1269.8;	DB 21;	Length 3217;
Best Local Similarity	81.18;	Pred. No. 0;		
Matches 1590;	Conservative 0;	Mismatches 277;	Indels 94;	Gaps 6;
QY 23	acttcctgcagcgtcgcgccccgagccgcgtctctcagcgtttgcgtgctgtcgc	82		
Db 1	acttcctgcggtcgcgcgccctgagcgcgtctcagcgtttgcgcgcgc--ggtgcgc	58		
QY 83	gtctgtgtgcgtccccctctctgag-----ccccgcctgcgcgccgcgc	132		
Db 59	gtctctctgcgtcccgcttcttgaccgcctcccccccgcgccgcgcgcgcgc	118		
QY 133	ttcgccctcgcgaactcgcgtcttccctcctcgtgtcgcgccttttctcgcgcgt	192		
Db 119	tcctccagcgcaactcgcgtcttctccctcctcgttgccttcttccctttttcc	178		
QY 193	caattgctcttcaactcctcctgcgcgcgcgccaagaccgcgcgcgcgcgcgcgc	252		
Db 179	ttcttcccttccctcctcgcgcgcgccaaccgcgcgcgcgcgcgcgcgcgcgc	238		
QY 253	ggaagcagcc-----	262		
Db 239	gcagcagccagattatlaaccacttaacctcagaaactgaacaagaacatgtgt	298		
QY 263	-----agaagtatagtcttaccattgagctcatgcaat	298		

D	b	299	cctggaacgcctcttctttaaaaaaagaagcataacccctactgtlagaactaaatgcact	358
Q	y	299	gtgtgtgaacctggaagaaaaaaccaatgtataagcccgltgaccctcactcgcgatgc	358
D	b	359	gtgcattgaaccttggaaaaaaaccaatgtataagcctgttgacccttactcgcgatgc	418
Q	y	359	gtccacctacagctatgcatgctgtgaggtgcctataccccccagatactttaccacit	418
D	b	419	gtccacctataactacacacatgagagaggtgtcttataccccccgagttactttaccacit	478
Q	y	419	tccagtlcccaaccttactactaccagaattgagctctccgttggcggacagcagttaatg	478
D	b	479	tccagtlcccaaccttactactlatacaagltggaacttctctgttggagacagcaattaatg	538
Q	y	479	gaaaggaagaatgagaccaccocgtgaaacacagatgccccctgccccgtgtgcgtgagactct	538
D	b	539	caaaggaagaagacacagacagcgtcgcgaacacagatgctgcgtgccaagcgttgagatcct	598
Q	y	539	gcagagtgtaacccccctgcacagaagaagtttgaggtataatlygaagagagacagaggaagaaa	598
D	b	599	gcagaaatgagccccctgcacagaagagcgtgtgaggtgaatlygaagagaaatccgaaagaaa	658
Q	y	599	cctcaataaatcggaaataagcccaagtgttgaatatgcgtgaagcggaaatttgcctgt	658
D	b	659	tctcaataaatctgaataaagtcgaagtgtttagatltgcacttaaacggaacttgcctgt	718
Q	y	659	gaattttgaggtgccccgggagagatlgccccaccacacatgaagaacttgtgaccaggt	718
D	b	719	gaattctgaggtgccccgggagagatlgccccaccacacatgaagaacttgtgaccaggt	778
Q	y	719	tctcagttggggaatttgttagggggaaggagaagggaanaagcaagaatctccaagaagaa	778
D	b	779	tctcagttggggaagttgttggggggaaggtgaagggaanaagcaagaatctccaagaagaaa	838
Q	y	779	tgcggccagggcgtgttctgtggcagccttagagagctgcccaccctccctgctgtgagcg	838
D	b	839	tgcgcgcataagctgttcttcttggaggagctlgaagaagtlacgcgccctgctgcagtgtgaacg	898
Q	y	839	agtgaaagcccaagaatcaagaagaanaagtlacgcccacctgc-----agacagccccga	892
D	b	899	agtgaaagccctagaatcaaaaaagaaaaaacaacccatagtlcaagccacagacaagcccaga	958
Q	y	893	ttatggccaagggatgtgaatcctatagtagacttgcacagatccagcagcgaaaaaagga	952
D	b	959	atatggccaaggggatcaatccgaattagccgactgccccagatccagcagcgaaaaaagga	1018
Q	y	953	gaaggaagccagagtlacatgctccttacagaacgaggtcttccacgttcgcagggagttgt	1012
D	b	1019	gaaggaagccagagtlacacgctctctcacagagcgaggtctccgcgcgcagggagttgt	1078
Q	y	1013	gatgcaggtaaaggttgggcatcacaactgcagaaagagtggtlaccataagaagtggc	1072
D	b	1079	gatgcaggtgaaggttggaaaaccacactgcagaaaggaacgggcaccaacaagaagtggc	1138
Q	y	1073	caagcgtaaatgctgtctgagacaatgctgagatcccttgggttcaaaagttccccagcgca	1132
D	b	1139	caagcgcaatgcaagccgagaacatgtgagatcccttgggttcaaaagttccccagcgca	1198
Q	y	1133	gcctgccaagccagcactcaaatcagaagaagaagactcagtaagaaacccagsgacgg	1192
D	b	1199	gcccaaccaaacccgcactcaagtctagagagagaagacaccataaagaaacccagsgatlg	1258
Q	y	1193	aaggaagaagtaacgttttttgaacctagccccctggggatgaanaatgaaactagtaacaagga	1252
D	b	1259	aaggaagaagtaaccttttttgaacctgtccttgggatatgaanaatggaactagtaataaga	1318
Q	y	1253	cgaagagttcagagatgccttacttagccatcacgacgtgcgcagctgtgaaattctcccat	1312
D	b	1319	ggatgagttcagagatgccttactataagtlcatcagcagctgcgtgtgaaattcttcccat	1378
Q	y	1313	ggtgcgggaagttgcccagcgtctgggggttagtcaaggaacacacacaaagaatttcac	1372
D	b	1379	ggtgcgggaaggttcgcccagcgtcttgaagagttagtcaaggaacacacacaaagaatttcac	1438

QY 1373 caggcagctcccaaatcctgcacaaagcaacgtaactgcatgatatagcccgagagtgtt 1432
|||||
Db 1439 caggcagctccgaatctctgcacaaagccacgtaactgcatgatatagcccgagagtgtt 1498
QY 1433 gtacggggcagctcgcgccacagccagaccattttaagaagtaacatctctcagagcca 1492
||| |||||
Db 1499 gtatggggcagctcgcgccacagccagaccattttaagaataacatctctcagagcca 1558
QY 1493 cgtaccccatgacctcgcacatagaccctctgagcaactgtactactcttccagagccca 1552
|||||
Db 1559 cgtaccccatgacctcgcacatagaccctctgagcaactgtactactcttccagagtcga 1618
QY 1553 gggatccaggttgaatacaagaatttcccaagaacaacaagaacgagtggtatctct 1612
|||||
Db 1619 gggatccaggttgaatacaagaacttcccaaaaacaacaagaacgaaattgtatctct 1678
QY 1613 catcaactgctcctcacagccgctctcgtcagtcatgtgcacggaagatgtgagtc 1672
|||||
Db 1679 tatcaattgctcctctcagccaccctcgtatcagcatgtgtatcggcaagatgtgagtc 1738
QY 1673 ctgtcatgatatggtgctgactgaacattttaagctgtctgtgagtttgaccacaacagag 1732
||| |||||
Db 1739 ctgcatgatatgctgtcgctgaacatcttaagttgctgtctgagttgaccacaacaaag 1798
QY 1733 cacagagatgcccaagaacacaggaatgaccagtttcagcgtgcggaggtgctgaacctt 1792
|||||
Db 1799 tacagagatgcccaagaacacaggaacccaatgtctgtgtgtyggaggtgctgaacctt 1858
QY 1793 ttctgcccacaaacattataaaa-cccaacatatatactgaaaaataactgagaactgctt 1851
|||||
Db 1859 ttctgcccatagaaccattataaaatcccaacatatatactgaaaaataactga-aactgctt 1917
QY 1852 tgaaaatttgaatatatctgataaactccagttggtgccaagaca 1892
|||||
Db 1918 tgaaaatttgaatttctgataacctccagttggtgccaagaga 1958

RESULT 4
X90786
ID X90786 standard; DNA: 3190 BP.
XX
AC X90786;
XX
DT 13-JAN-2000 (first entry)
XX
DE DNA encoding human stauflen protein.
XX
KW Human.staufen protein; hStau; protein lysate; lung; kidney; testis;
KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;
KW telomerase assembly; maturation; transport; regulation; interaction;
KW inhibitory agent; antibody; immunoassay; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 345..1835
FT /*tag= a
FT /product= "Human stauflen protein"
FT /note= "Binds to telomerase RNA"
FT 3172..3177
FT /*tag= b
FT
FT polyA_signal
FT
FT
PN WO951255-A1.
XX
XX 14-OCT-1999.
PD
XX
PF 06-APR-1999; 99WO-US07533.
XX
XX 06-APR-1998; 98US-0080783.
PR
XX
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX

PI Greider CW, Le S;
XX
XX WPI: 1999-620168/53.
DR P-PSDB; Y26335.
XX
PT Human stauflen polypeptide useful in methods for identifying telomerase
PT inhibitors
PS Claim 13; Page 41-42; 50pp; English.
XX
XX The present sequence is an isolated DNA encoding human stauflen (hStau)
CC protein. The hStau protein was detected in protein lysates from lung,
CC kidney, testis and ovary, but not in brain and heart. It binds
CC specifically to human telomerase RNA and can complex with the
CC telomerase catalytic subunit. The hStau protein may play a role in
CC telomerase assembly, maturation, transport and regulation. Interaction
CC between hStau and telomerase can be inhibited by administering inhibitory
CC agents to the cell. These telomerase inhibitors can be identified using
CC hStau protein. Antibodies generated against hStau can be used in various
CC immunoassays.
XX
SQ Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;

Query Match 44.0%; Score 1256.6; DB 20; Length 3190;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1420; Conservative 0; Mismatches 209; Indels 8; Gaps 3;

QY 263 agaaagataagcttaccattgagctcaatgacactgtgtgaaactggaagaaaacc 322
|||||
Db 305 agaaagcalaaccctactagttagaactaaatgacactgtgcatgaaacttggaaaaaac 364
QY 323 aatgtataagcccgtygacccctcactctcgatgcagttccaactacagctatgcatgcg 382
|||||
Db 365 aatgtataagcctgttgaccccttactctcgatgcggtccaactataactacaacatgag 424
QY 383 tggaggtgcctatccccagatacttttaccatttccaagttccaacttacttacc 442
|||||
Db 425 aggagtgcttataccccgagtagtactttaccatttccaagttccaacttacttacc 484
QY 443 agttgagctctccgttggcgacacagtttlaatggaagaaagaaatgagaccaccgt 502
||| |||
Db 485 agtgaacttctgttggagacagcaatttaatggaagaaagaaagacagagctgc 544
QY 503 gaacacagatgcccttgcctgctgagagactctgcagagtgaaacccctgcagaagag 562
|||||
Db 545 gaacacagatcgtctgcacaaagcgttgagatcctgcagaatgagccctgcagaagag 604
QY 563 gttgaggttaaatggaagaaagcagaagaaacccaataatcgaaataagcca 622
||| |||||
Db 605 gctggaaggttaatggaagaaatccgaagaaataatccaataatctgaataagtc 664
QY 623 agtgtttgaattgctgtaagcggaaatttgctgtgaatttgaggtggtgcccggagag 682
|||||
Db 665 agtgtttgaattgctgtaagcggaaatttgctgtgaatttgaggtggtgcccggagag 724
QY 683 tggccaccacacatgaagaacttctgaccaaaggttccaagtttggaatttgtagggga 742
|||||
Db 725 tggccaccacacatgaagaacttctgaccaaaggttccaagtttggaatttgtagggga 784
QY 743 aggaagaggaagaaagcagaagatctccaagaagaatgcggccaggtgttctggaagca 802
||| |||||
Db 785 agtgaagggaaaagcagaagattccaagaaaatgcccacatagctgttcttgagga 844
QY 803 gcttagaggtctgccaccctccctgctgtgagcgagtgaaagccagaatcaagaagaa 862
||| |||
Db 845 gctgaagaaagttaccgcccctgctgcagttgaaacgagtaagcctagaatcaaaaagaa 904
QY 863 aagtcaagccacactgca-----agacagcccccagattatggccaaggtatgatactat 916
||| |||||
Db 905 aacaaaaccatagttcaagccacagaacaaagccagaatatggtgccaaggtatcaatccgat 964
QY 917 tagtaacttgcacagatccagcagggcaaaaaaggaaggaagcagagtagtactgtcct 976

Db 965 tagccgactgcccagatccagcagcgcaaaaagagaagagagccaggtacaagctcct 1024
QY 977 tacagaacgaggtcttccacgctgcagggagttgtatgcaggtaaaggttggcatca 1036
Db 1025 cacagagcgagcctcccgccgcgcagggagttgtatgcaggtgaaggttgaacca 1084
QY 1037 cactgcagaagaggtgtaccaataagaagtggtccaaagcgttaatgtctgtagaacat 1096
Db 1085 cactgcagaaggaacgggcaccacaagaagtggtccaagcgcaatgcagccgagaaacat 1144
QY 1097 gctggagatcctgggttccaaagttcccgagcgagccttgccaagccagctcaaatc 1156
Db 1145 gctggagatcctgttccaagttcccgagcgagcctccaacccgcagctcaagtc 1204
QY 1157 agaagagaagactccagtaaaagaaccagagagcggaagaaagtaacgtttttgaacc 1216
Db 1205 agagagagaagacccataaagaacacaggggagtgaagaaagtaacctttttgaacc 1264
QY 1217 tagccctgggagatgaaaaatggaactagtaacaaggaagcgaggttcagatgccttatct 1276
Db 1265 tggctctgggagatgaaaaatggagactagtaataagaagatgagttcagatgccttatct 1324
QY 1277 tagccatcagcagctgcagcagctgtgaattctcccatgtgtccggaagttgcccaggtgt 1336
Db 1325 aagtcatcagcagctgcctgtctggaattcttcccatgtgtccggaagttgcccaggtgt 1384
QY 1337 cggggtagtcaaggaacacacacacaaagatttcaccagggcgagctccaatcctgcacaa 1396
Db 1385 aggagtagtcaaggaacacacacacaaagatttcaccagggcgagctccaatcctgcacaa 1444
QY 1397 ggcaacggttaactgcacatgatagcccagagagtggtgtacgggggcacctcgcccacagc 1456
Db 1445 ggccacggttaactgcacatgatagcccagagagtggtgtacgggggcacctcgcccacagc 1504
QY 1457 cgagaccattttaagagtaacatctcttcagggccaagtaaccatggaactcgcaatag 1516
Db 1505 cgagaccattttaagagataacatctcttcagggccaagtaaccatggaactcgcaatag 1564
QY 1517 accctctgagcaactgttactactcttccagagcccagggatccaggttgaataacaaga 1576
Db 1565 accctctgagcaactgttactactcttccagagtcagggatccaggttgaataacaaga 1624
QY 1577 ttctcccaagaacaagaacgagtggtgtatctctcatcaactgtctctcaacagccgcc 1636
Db 1625 ctccccaagaacaagaacgagtggtgtatctctcatcaactgtctctcaacagccacc 1684
QY 1637 tctcgtcagtcagtcgacggaagagatgtgagtcctgtcatatgatatgctgcactgaa 1696
Db 1685 tctgatacgccatgtatcggaagagatgtgagtcctgtcatatgatatgctgcactgaa 1744
QY 1697 catlttaagctgtctgtgagtttgaccacaagagcacagagatgccaagaacaggaaa 1756
Db 1745 catlttaagctgtctgtgagtttgaccacaagaagtaacagatgccaagaacaggaaa 1804
QY 1757 tggaccagtttcagcggtcggggaggtgtgtgaaccttttctggccacaacacattataaa 1816
Db 1805 cggaccacatgtctgtgtggtggaggtgtgtgaaccttttctggccatgaaccattataaa 1864
QY 1817 -cccacaatatatactgaaaaatactggaagactgcttgaanaatttgaatatctgataac 1875
Db 1865 tcccaacatatatactgaaaaatactga-aactgtcttgaaaaatttgaatttctgatacc 1923
QY 1876 tccagtgggccaagaca 1892
Db 1924 tccagtgggccgagaga 1940

RESULT 5
ID 293322 standard; cDNA; 3506 BP.
XX
AC 293322;

XX 04-JUL-2000 (first entry)
XX Human staufen cDNA.
XX Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 409..2142
FT /*tag= a
FT /product= Staufen
XX CA2238656-A1.
PD 22-NOV-1999.
XX 22-MAY-1998; 98CA-2238656.
PF 22-MAY-1998; 98CA-2238656.
PR 22-MAY-1998; 98CA-2238656.
XX (UYMO-) UNIV MONTREAL.
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L;
PI Mouland AJ;
XX WPI; 2000-246924/22.
DR P-PSDB; Y83023.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hSP (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

Query Match 44.0%; Score 1256; DB 21; Length 3506;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1420; Conservative 0; Mismatches 210; Indels 8; Gaps 3;

QY 262 cagaagatatagcttaccatgagctcaatgacatgtgtgaaactggaagaaaac 321
Db 611 cagaagataaacccctactgtgaactaaatgacatgtgcatgaacttgaaaaaaac 670
QY 322 caatgtataagcccgtygacctcaactctcgatgcagtcacaccatcacagtgc 381
Db 671 caatgtataagcctgttgaccttactctcgatgcagtcacaccataactacaacatga 730
QY 382 gtagaggtgcctatcccccgataacttttaccattccagttccacacttacttacc 441
Db 731 gagaggtgtctatcccccgagtaacttttaccattccagttccacacttacttacc 790

QY 442 aagltgagctctccgtggygcgacagcagtttaatgagaaagaaagatgagaccaccg 501
Db 791 aagltgaaacttctgtggyagagacagcaatttaatgcaaaagaaagacaagacagctg 850
QY 502 tgaacacagatgccccctgccccgtgtgcgtgagagactctgcagagtgaacccccctgcagaa 561
Db 851 cgaacacagatgtctgtcccaagcgttgtgagatcctgcagaaatgagccccctgcagaga 910
QY 562 ggttgaggttaaatggaagagaagacagagagaagaaacccaataaatcgaataaagcc 621
Db 911 ggttgaggtgtaatggaagagaatccgaagagaanaatccaataaatcgaataaagtcc 970
QY 622 aagtgttgaatgtgcgtgaagcaggaattgtgcctgtgtaattttgaggtgccccggaga 681
Db 971 aagtgttgaatgtgcaacttaaacggaacttgccctgtgtaatttcgaaggtgccccggaga 1030
QY 682 gtgccccacacacatgaagaactttgtgacacaggtttcaagttggygaatttgtaggg 741
Db 1031 gtgccccacccccacatgaagaactttgtgaccaaagtttcggttggygaatttgtggygg 1090
QY 742 aaggaagaaggaaagcaagaagatctccaagaagaatgcgcgcaggtctgtctgagc 801
Db 1091 aaggtgaaggaaagcaagaagattccaagaaaaatgcgcgcatagtctgtcttgag 1150
QY 802 agcttagagagctgcacacccccctcctgtctgtgagcgagtgaaagcccaagaatcaaga 861
Db 1151 agctgaagaagttaccgccccctgtcgtcagtgaaagtaaaagcctagaatacaaaaa 1210
QY 862 aaagtcagccccactgca-----agacagccccggattatgccaagggatgatccta 915
Db 1211 aaacaaaccatagtcacagccacagacaagccccagaatatgcccaggggatcaatccga 1270
QY 916 ttagttagcttgcacagatccagcagcacaanaaggaagagccagagtaacatgctcc 975
Db 1271 ttagccgactggtcccaagatccagcagcacaanaaggaagagccagagtaacatgctcc 1330
QY 976 ttacagaacgaggtcttccacgtgcgaggaagtttgtatgcaggtaaagttggygcatc 1035
Db 1331 tcacagagcgaagcctccgcgcgcgcaggaagtttgtatgcaggtgaagttggaacc 1390
QY 1036 acactgcagaagagtggtgtaccaataagaagtgycacaagcgtaatgctgtgagaaca 1095
Db 1391 acactgcagaagagcggcaccacaagaagtgycacaagcgaatgcagccgagaaca 1450
QY 1096 tgctgagatcctggyggttcaaaattccccagggcgagcctgcgaagccagcactcaaat 1155
Db 1451 tgctgagatcctggtttccaagttcccgagcgcgagccaccaaaacccgacactcaagt 1510
QY 1156 cagaagagaagactccagtaaagaaacccaagagacggaaggaagtaacgttttttgaa 1215
Db 1511 cagaagagaagacacccataaagaaaccaaagggtggaagaaagtaacccctttttgaa 1570
QY 1216 ctagccctggyggtgaaatggaactagtaacaaggaagagagttcagatgacctatc 1275
Db 1571 ctggtctggygtgaaatggtgactagtaataaagagatgagttcagatgacctatc 1630
QY 1276 ttagccatcagcagctgcagctggaattctcccatgtgtgcggaagttgcccaggtg 1335
Db 1631 taagtcatcagcagctgctgtgtaattctcccatgtgtgcggaagttgcccaggtg 1690
QY 1336 tcggygttagtcaaggaacacacacacaaagatttcacacagggcagctccaatcctgcca 1395
Db 1691 taggaagttagcaaggaacatcacacacaaagattttaccaaagggcagctccgaatccgcca 1750
QY 1396 aggcacaagtgtaactgcatgatagcccgaagagttgtgtacggyggcacctgcgccacag 1455
Db 1751 aggcacaagtgtaactgcatgatagcccgaagagttgtgtacggyggcacctgcgccacag 1810
QY 1456 ccgaagaccatttaaaagagtaacatcttccagggccacgtaaccacatgagacctgcgacata 1515
Db 1811 ccgaagaccatttaaaagataacatcttccagggccacgtaaccacatgagacctgcgacata 1870
QY 1516 gaccctctgagcaactgtactaccttccagagcccaaggtatccaggttgatacaaaag 1575

Db 1871 gaccctctgagcaacgagactatcttccagaggtccagggatccaggttgatacaaaag 1930
QY 1576 atttcccaagaacaacaagaagagtggtatctctcatcaactgtctcctcacagcgc 1635
Db 1931 acttcccaaaaaacaacaagaagcaatgtatctctcatcaactgtctcctcacagc 1990
QY 1636 ctctcgtcagtcagtcacgcgcgaagagatgtgagtcctgtcatgatatgtgctgactga 1695
Db 1991 ctctgatcagccatggtatcgcgaagagatgtgagtcctgtccatgatatgtgctgactga 2050
QY 1696 acattttaaagctgtctgtagttggaaccaacagagacagagatgccaagaagagaa 1755
Db 2051 acatcttaaagttgtctgtagttggaaccaacaaagtlacagagatgccaagaagagaa 2110
QY 1756 atgaccagtttcagcggtgcggaaggtgtcgaaccttttctgcccacaacacataaa 1815
Db 2111 acggaaccaatgtctgtgtggyaggtgtcgaaccttttctgcccataaaccattataa 2170
QY 1816 a-cccaacatataactgaaataactgaggaactgtttgaaatttggaaatatctgataa 1874
Db 2171 atcccaacatataactgaaataactga-aactgcttggaaatttggaaatttctgtac 2229
QY 1875 ctccagtggyccaagaca 1892
Db 2230 ctccagtggyccgagaga 2247

RESULT 6
Z93327
ID Z93327 standard; cDNA; 3260 BP.
XX
AC Z93327;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; ; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 409..1311
FT /*tag= a
FT /product= Stauflen protein
XX
CA2238656-A1.
XX
PN 22-NOV-1999.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;
PI Moulard AJ;
XX
DR WPI: 2000-246924/22.
DR P-PSDB; Y83108.
XX
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
XX
PS Claim 4; Figure 1b; 96pp; English.
CC Stauflen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA

CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC stauften with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian stauften
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/mSP (murine SP) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauften proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC Z93332.

Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

Query Match	43.4%	Score 1239;	DB 21;	length 3260;
Best Local Similarity	86.5%;	Pred. No. 0;		
Matches 1416; Conservative	0;	Mismatches 210;	Indels 11;	Gaps 4;

[illegible]

Db	1086	cacagagcgaagcctccgcgcgcgacgagagtcttgatgcaggtgaaggttgaaacca	1145
QY	1037	cactgcagaagagatgtggttaaccaataagaaggtgcccgaagcgtaatgtctgtgaaacat	1096
Db	1146	cactgcagaagaagaaacggtccaccaacaagaaggtgccaagcgcaatgcaagccggaacat	1205
QY	1097	gctgagatcctggggttcaaaagttcccagggcgagcctgtccaagccagcactcaaatc	1156
Db	1206	gctggaatccttgtttcaaaagttccgcagcgcagcccaacaaaccgcactcaagtc	1265
QY	1157	agaaagaaagactccagtaaaagaaaccaggaagacggaagaaagtaacgtttttgaacc	1216
Db	1266	agagagaagacacccataaagaaacacgaggatgtaagaagaaagtaaccttltgaacc	1325
QY	1217	tagccctggggtatgaaaatgtgaactagtaacaagagcaggagtlcagatgctcatct	1276
Db	1326	tgctctggggtatgaanaatgaggactagtaataaaggagatgagtlcagatgcctatct	1385
QY	1277	tagccatcagcagctgtccagctgtgaattctcccacgtgtgccggaagtlgtcccagctgt	1336
Db	1386	aagtcatacagcagctgtccctgtgtgaattcttcccacgtgtgccgaggtgcgccagctgt	1445
QY	1337	cggggttagtcaagagacaccacaccaaagatttcaccaaaggcagctccaatcctgtccaa	1396
Db	1446	aggaagttagtcaaagacatcacaccaagaatttaccagggcagctccgaatcctgtccaa	1505
QY	1397	ggcaacggttaactgcccatagtacccgagagttgttgaacgggggcacctgcgccacagc	1456
Db	1506	ggccacggttaactgcccatagtacccgagagttgttgaacgggggcacctgcgccacagc	1565
QY	1457	cgagaccattttaagagtaacatctcttcagggccagctlaccccatgacctgcagctag	1516
Db	1566	cgagaccattttaagagtaacatctcttcagggccagctlaccccatgacctgcagcag	1625
QY	1517	accctctgagcaactgttactaccttccagagcccagggattccaggttgaatacaaaaga	1576
Db	1626	accctctgagcaactgttactaccttccagagctccagggattccaggttgaatacaaaaga	1685
QY	1577	ttttcccaagaacaacaagaacgagtggtatctctcatcaactgtctctcacagccgc	1636
Db	1686	cttcccaagaacaacaagaacgaaattgtatctctcatcaactgtctctcacagccacc	1745
QY	1637	tctcgtcagtcagtcgcatcgccaggaagtgtgagtcctgtcatgatgctgctgaactgaa	1696
Db	1746	tctgatcagcagtcgtatcgccaagatgtgtgagtcctgtcagtcagatgctgcgtgaa	1805
QY	1697	catttaagctgtctgtctgagttggaaccaacagagcagagatgccaagaacaggaaa	1756
Db	1806	catcttaagttgtctgtctgagttggaaccaacaagaatgagatgccaagaacaggaaa	1865
QY	1757	tgagccagtttcagcgtgcgggaggtgtctgaaccttctgtgccacaacccattataaaa	1816
Db	1866	cggagccaatgtctgtgtgtggaaggtgtctgaaccttctgtgccatgaaccattataaaa	1925
QY	1817	-cccaacatatatactgaaataactgagaactgctltgaaaaatttgaatatctgataac	1875
Db	1926	tcaccaacatatatactgaaataactgga-aactgctltgaaaaatttgaaatctgataacc	1984
QY	1876	tcacagltggccaagaca 1892	
Db	1985	tcacagltggccaagaca 2001	

RESULT	7
X39991	
ID	X39991 standard; DNA; 1127 BP.
XX	
AC	X39991;
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Prostate cancer associated gene
XX	

KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
PI Pfrendschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI: 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT Isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 625-626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;

Db 241 agaacttgtgaccgaaggttcggttgaggagttgtggggaaggtgaaggaaagca 300
QY 760 agaagatctccaagaagaatcgcgccaggctgtcttgagcagcttagaggtccac 819
Db 301 agaagattccaagaaaaatgcgcctagctgtcttgagagcttgaaagattaccgc 360
QY 820 ccctccctgtgtgagcgagtgaaagcccaaatcaagaagaagaagtcagccacctgca 879
Db 361 ccctgcctgcagttgaacgagtaaaagcctagaatcaaaaagaaaaaccacatagta 420
QY 880 -----agacagcccgatlatgcccaggatgaatcctattagtagactgacaga 933
Db 421 agccacagacaagcccgagaatatgcccagggtatcaatccgattagccgactgcccaga 480
QY 934 tccagcagcaaaaaaggaaggaagcacaagttacatgctccttacagaacgaggtcttc 993
Db 481 tccagcagcaaaaaaggaaggaagcacaagttacacgctcctcacagaagcgagcctcc 540
QY 994 cagctcgagggagttgtgatgcaagtaagggttggcatcacactgcagaagagtg 1053
Db 541 cgcgcgcgaggtgtgtgatgcaagttgaaggtltgaaaccacactgcagaaggaacgg 600
QY 1054 gtaccaataagaaggtggccaagcgtaatgtcgtgtgagaacatgctggaagatcctgggt 1113
Db 601 gcaccaacaagaaggtggccaagcgtaatgtcgcgcgagacatgctggagatccttggt 660
QY 1114 tcaaagttcccgagcgcagcctgcccaggccagcactcaatcagaagaagaagactccag 1173
Db 661 tcaaagttcccgagcgcagcctgcccaggccagcactcaatcagaagaagaagactccag 1233
QY 1174 taaagaaccaggaagcgaaggaagaagtaagtttttgaacctgaagccttgggatgtaa 1293
Db 721 taaagaaccaggaagcgaaggaagaagtaagtttttgaacctgaagccttgggatgtaa 780
QY 1234 atggaactagtaacaagcagcaggaagttcagatgcttatcttaccatcagcagctgc 1293
Db 781 atggactagtaataaagaagatgagttcagatgcttatcttaccatcagcagctgc 840
QY 1294 cagctggaattctcccatgtgcccgaaggtgcccagcgtgcgggttagtcaagagc 1353
Db 841 ctgctggaattctcccatgtgcccgaaggtgcccagcgtgcgggttagtcaagagc 900
QY 1354 accacaccaagaatttcaccagcgcagctccaatcctgcgaagcgaacgtaactgcca 1413
Db 901 atcacaccaagaatttcaccagcgcagctccaatcctgcgaagcgaacgtaactgcca 960
QY 1414 tgatagcccgagagttgtgtlacgggggcacctgcgccacag-ccgagaccattttaa 1472
Db 961 tgatagcccgagagttgtgtlacgggggcacctgcgccacagcccgagaccattttaa 1020
QY 1473 agtaacatctcttaaggccaagtcaccccatggaactgcgaactagaccctctgagcaact 1532
Db 1021 aataacatctcttaaggccaagtcaccccatggaactgcgaactagaccctctgagcaact 1080
QY 1533 tactaccttcagagcccaagcaggtatccaggttgaatacaaaagattt 1579
Db 1081 gactatcttncagagtcacaggtatncaggttgaatacaaaagattt 1127
RESULT 8
X39992
ID X39992 standard; DNA: 773 BP.
XX
AC X39992;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX

OS	Homo sapiens.
XX	
PN	W09904265-A2.
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998; 98WO-US14679.
XX	
PR	22-JUN-1998; 98US-0102322.
PR	17-JUL-1997; 97US-0896164.
PR	10-OCT-1997; 97US-0061599.
PR	10-OCT-1997; 97US-0061765.
PR	10-OCT-1997; 97US-0948705.
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDM-) LUDWIG INST CANCER RES.
XX	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI	Tureci O;
XX	
DR	WPI; 1999-132448/11.
XX	
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Claim 67; Page 626; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
SQ	Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other:
Query Match	20.6%; Score 588.8; DB 20; Length 773;
Best Local Similarity	86.8%; Pred. No. 4.6e-166;
Matches 671; Conservative	0; Mismatches 95; Indels 7; Gaps 2,
OY	814 tgcacccctccctgtgtgagcgagtgaagcccagaatcaagaagaaagtcagcca 873
Dd	1 taccgccctgtcgtcaglttgacgagtaaaagcctaagaatcaaaaagaacaaccoca 60
OY	874 cctyca-----agacagccccgatattatggccaaggatgaatcctttagtactg 927
Dd	61 tagtcaagcacacagaagcccagaatalatgcccaggygatcaatcogattagccgactg 120
OY	928 cacagatccagcaggcacaaaaagagagaagcagcagatcatatgcttcttacagaacgag 987
Dd	121 ccagatccagcaggcacaaaaagagagaagcagcagatcatatgcttcttacagaacgag 180
OY	988 gtcttccagctgcagggagtttgtatgtcaggttaaaggttgycatcacactgcagaag 1047
Dd	181 gctctccgcgcgcgcagggagtttgtatgtcaggttaaaggttgycatcacactgcagaag 240
OY	1048 gaatgggtaccaataagaagtggccaagcgtaatgtcgtctgaagaacatgtcygagatcc 1107
Dd	241 gaacgggcaccaacaagaagtggccaagcgcaatgtcagccgagaaacatgtcygagatcc 300
OY	1108 tggggttcaaagttccccagcgccagcctgccaagccagcagcatcataatcagaagaga 1167

Db	301	ttgttttcaaaagtcctccgcagcgccagcccaaaaccgcactcaagtcaagtagagaaga	360
Qy	1168	ctccagtaaagaacaccgagagacggaaggaagttaacgttttttgaacctagccctgggg	1227
Db	361	caccataaagaacaccaggggatggaagaagaagtaaccttttttgaacctggtctgggg	420
Qy	1228	atgaaaatggaactagtaacaagagcagagaggttcagatgctctatctttagccatcagc	1287
Db	421	atgaaaatggaactagtaacaagagatgagatgttcagatgctctatctttagccatcagc	480
Qy	1288	agctgcagctggaattctcccatggtgcgcggaagttgccaggtctgcgggttagtc	1347
Db	481	agctgcctgtggaattcttcccatggtgcgcggaagttgccaggtctgtagagtagtc	540
Qy	1348	aaggaaccacacccaagaatttcaccagggcagctccaatctctgccaaggaacggtaa	1407
Db	541	aaggaaccacacccaagaatttcaccagggcagctccaatctctgccaaggaacggtaa	600
Qy	1408	ctgcacatgatagcgccgagagtggtgtgacgggggcacctgcgccacag-ccgagaccatt	1466
Db	601	ctgcacatgatagcgccgagagtggtgtgacgggggcacctgcgccacagcccgagaccatt	660
Qy	1467	ttaaagagtaacatctcttcagggccacgtaccatgtgacctgcgcatagaccctctgag	1526
Db	661	ttaaagaataacatctcttcagggccacgtaccatgtgacctctcagcagagaccctntgag	720
Qy	1527	caactgtactactcttcacagagccagggatccaggttgtaatacaagaattt	1579
Db	721	caactgtactactcttncagaggtccagggatcncaggttgtaatacaagaactt	773
RESULT	9		
X40002	ID	X40002 standard; DNA; 946 BP.	
XX	AC	X40002;	
XX	DT	02-JUL-1999 (first entry)	
DE	XX	Prostate cancer associated gene.	
KW	XX	Cancer associated antigen; diagnosis; research; treatment; human;	
KW	XX	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KW	XX	prostate cancer; ss.	
OS	XX	Homo sapiens.	
PN	XX	WO9904265-A2.	
PD	XX	28-JAN-1999.	
PF	XX	15-JUL-1998; 98WO-US14679.	
PR	XX	22-JUN-1998; 98US-0102322.	
PR	XX	17-JUL-1997; 97US-0896164.	
PR	XX	10-OCT-1997; 97US-0061599.	
PR	XX	10-OCT-1997; 97US-0061765.	
PR	XX	10-OCT-1997; 97US-0948705.	
PR	XX	11-OCT-1997; 97GB-0021697.	
PA	XX	(LUDW-) LUDWIG INST CANCER RES.	
PI	XX	Chen Y, Gout I, Gure A, O'Hare M, Odata Y, Old LJ;	
PI	XX	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;	
PI	XX	Tureci O;	
PT	XX	WPI; 1999-132448/11.	
PT	XX	New isolated cancer associated nucleic acids and polypeptides -	
PT	XX	isolated using sera from cancer patients, used to develop products	
PT	XX	for the diagnosis, monitoring or treatment of cancers	
PS	XX	Claim 67; Page 629-630; 787pp; English.	

|||||
Db 236 TTCAGATGCTTATCTAAGTCATCAGCAGCTGCCTGCTGGAATTCTTCCCATGTGCC 177
QY 1320 gaatgtcccaaggctgtcgggtatgtcaagagacacacacaaagatttcaccaggga 1379
|||
Db 176 GAGTGTGCGCCAGGCTGTAGAGTTAGTCAAGACATCACCAAGATTTTACAGGGCA 117
QY 1380 gctccaatcctgccaagcagtaactgcatgatagcccgagagattgtgtacgg 1439
|||||
Db 116 GCTCCGAATCCTGCCAAGGCCGCTAACTGCATGATAGCCCGAGAGTTGTATGGG 57
QY 1440 ggcacctgcgccacagccgagaccatttaagagtaacatctctcagycacgt 1495
|||||
Db 56 GGCACCTGCGCCACAGCCGAGACCATTTTAAAGAATAACATCTCTCAGGCCACGT 1

RESULT 12
280269
ID 280269 standard; cDNA; 727 BP.

XX 280269;
AC
XX
DT 07-APR-2000 (first entry)

XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.

DE Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytosstatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.

XX Homo sapiens.

OS
XX
PN WO9964576-A2.

XX 16-DEC-1999.

XX 09-JUN-1999; 99WO-IB01062.

XX 10-JUN-1998; 98US-0088801.

XX (FARB) BAYER CORP.

XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;

XX WPI; 2000-087220/07.

XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -

PS Claim 15; Page 273; 469pp; English.

XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.

XX
SQ Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;

Query Match 10.7%; Score 306.4; DB 21; Length 727;
Best Local Similarity 83.8%; Pred. No. 1.6e-81;
Matches 357; Conservative 0; Mismatches 68; Indels 1; Gaps 1;

QY 403 gatacttaccatttccagctccacccttactctaccaagttgagctctcgtggcg 462
|
Db 1 gtlacttlaaccatttccagtlccacccttacttlatcaagtggaacttctcgtggag 60
QY 463 gacagcagtttaatgtggaaggaagatgagaccaccgtgaaacacagatgccctggcc 522
|||||
Db 61 gacagcaatttaatgtgcaaaaggaagacagacagcgtcgaacaacagatgctgtcca 120
QY 523 gtgcgtcgaagactctgcagagtgaaacccctgcagaaagtttgaggtataatggaag 582
|||
Db 121 aagcgttgaagatcctgcagaatgagccctgcagagaggtctgaggtgaatggaag 180
QY 583 aagcagaagaaagaaaacctaataatcggaataaagccaagtgttgaatctgcgtga 642
|||
Db 181 aatccgaagaaagaaaatctcaataaatctgaataaagtaagtglttgagatgcaactta 240
QY 643 agcgaattgcctgtgaattttgaggtgcccggagagtgccccacacacatgaaga 702
|
Db 241 aacggaactgtcgtgtgaatttcgaggtgcccggagagtgccccacacacatgaaga 300
QY 703 acttgtgaccaggttctcagttggtgggaattgtlaggggaaggaagggaaagcaaga 762
|||||
Db 301 acttgtgaccaggttctcgtgtgggagttgttggtgggaaggtgaagggaaagcaaga 360
QY 763 agatctccaagaagaatg-cggccaaggctgttctctgagcagcttagagagctgccacc 821
|||||
Db 361 agatttcaagaagaaaatgtcccgccatagctgntcttgaggagctgaagaagatccgnc 420
QY 822 ctccct 827
|||
Db 421 ctggt 426

RESULT 13
217757
ID 217757 standard; cDNA; 300 BP.

XX 217757;

XX 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:5230.

DE Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

OS
XX
PN WO9938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 28-JAN-1998; 98US-0072910.

XX 24-FEB-1998; 98US-0075954.

XX 31-MAR-1998; 98US-0080114.

XX 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Gliese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

```
OY 318 caaccctccacctcccaagt--cagcgatttcccaactcagccctcccgataagctgagat 375
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205982 CAACCTCTACCTCCAGGCTCAAGCAATCTCTGCTCAGCCCTCCCGAGTAGCTGGAT 205923
OY 376 tacagagttattaaccac 394
      ||||| | | |||||
Db 205922 TACAGCATGCATCACCAC 205904

RESULT 4
US-07-906-871-15
; Sequence 15, Application US/07906871
; Patent No. 5340739
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Avraham, Shalom
; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W., Suite 300
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/906,871
; FILING DATE: 19920103
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/816,289
; FILING DATE: 03 JAN 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/635,544
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US89/03051
; FILING DATE: 13-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/224,035
; FILING DATE: 13-JUL-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0627.2830004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)833-7533
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17327 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: exon
; LOCATION: 621..753
; FEATURE:
; NAME/KEY: intron
; LOCATION: 754..9596
; FEATURE:
; NAME/KEY: exon
; LOCATION: 9597..9744
; FEATURE:
```

```
; NAME/KEY: intron
; LOCATION: 9745..16396
; FEATURE:
; NAME/KEY: exon
; LOCATION: 16397..17327
; US-07-906-871-15

Query Match 2.6%; Score 86.8; DB 1; Length 17327;
Best Local Similarity 82.4%; Pred. No. 1.7e-12;
Matches 112; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

OY 261 tggagtttgctctgtgcgccagcgctgagtgagtgagtgatctcggtcactgcaa 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15481 TGGAGTTTGTCTGTGTACCCAGGCTGAGTGCAGTGGCGGCGATCTCGGCTCAGTGCAA 15540
OY 321 cctccacctcccaagt--cagcgatttcccaactcagccctcccgataagctgagattac 378
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15541 CCTCCGCCCTCCTGGGTCAAGCGATTCTCTCAGCTCAGCCCTTCCGAGTAGCTGGGATTAC 15600
OY 379 agagtttattaaccac 394
      || | |||||
Db 15601 AGGCATGCACCACCAC 15616

RESULT 5
US-08-370-319C-12/C
; Sequence 12, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coulle, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN ANTIGEN PRESENTED TO AT LEAST ONE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,319C
; FILING DATE: 10-JANUARY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/272,351
; FILING DATE: 8-JULY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/032,978
; FILING DATE: 18-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5856091man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5377.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
```


QY	318	caacctccacctcccaggt--cagcgattttccaccatcagctcccgataagtctgagt	375
Db	141312	CAACCTCTGCGCTCCCGGTTCAGACAATTCTCCTGCTCAGCCTCCGGAGTAGCTGGAT	141253
QY	376	tacag	380
Db	141252	TACAG	141248

```

RESULT 10
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
;
GENERAL INFORMATION:
;
APPLICANT: Pan, Yang
;
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
;
FILE REFERENCE: 09404/052001
;
CURRENT APPLICATION NUMBER: US/09/128,155
;
CURRENT FILING DATE: 1998-08-03
;
EARLIER APPLICATION NUMBER: US 60/091,650
;
EARLIER FILING DATE: 1998-07-02
;
EARLIER APPLICATION NUMBER: US 60/054,646
;
EARLIER FILING DATE: 1997-08-04
;
NUMBER OF SEQ ID NOS: 18
;
SOFTWARE: FastSeq for Windows Version 3.0
;
SEQ ID NO 17
;
LENGTH: 176373
;
TYPE: DNA
;
ORGANISM: Homo sapiens
;
FEATURE:
;
NAME/KEY: misc_feature
;
LOCATION: (1)..(176373)
;
OTHER INFORMATION: n = A,T,C or G
;
US-09-128-155-17

```

Query Match	2.5%;	Score 82.2;	DB 3;	Length 176373;
Best Local Similarity	84.0%;	Pred. No. 9.2e-11;		
Matches 105;	Conservative	0;	Mismatches 18;	Indels 2;
				Gaps 1;

Qy	258	aggtgagatttgcctcttctgcgcccaagctgagtcagtcgtgatactcgcgtcaatg	317
Db	25365	AGATGGAGTTTCACTCTTGTGGCCAGGCTGGAGTGCATGGTGTGATCTCGGCTCACTG	25306
Qy	318	caacctccacctcccaagt--cagcgatttcccaacttcagcctcccgataagctgagat	
Db	25305	CAACCTCTGCTCCCGGGGTTCAAGCAATTTCTCTGCCTCAGCCTCCCGAGTAGCTGGAGT	25246

QY	376	tacag	380
Db	25245	TACAG	25241

RESULT 11
 US-09-014-969-1
 ; Sequence 1, Application US/09014969
 ; Patent No. 5965397
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: Lavallee, Edward R.
 ; APPLICANT: Racie, Lisa A.
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Agostino, Michael J.
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 ; TITLE OF INVENTION: ENCODING THEM
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive

```

1 CITY: Cambridge
2 STATE: MA
3 COUNTRY: U.S.A.
4 ZIP: 02140
5
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: Patent In Release #1.0, Version #1.30
11
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/09/014,969
14
15 FILING DATE:
16
17 CLASSIFICATION:
18
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Sprunger, Suzanne A.
21 REGISTRATION NUMBER: 41,323
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: (617) 498-8284
24 TELEFAX: (617) 876-5851
25
26 INFORMATION FOR SEQ ID NO: 1:
27
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 2509 base pairs
30 TYPE: nucleic acid
31 STRANDEDNESS: double
32 TOPOLOGY: linear
33
34 MOLECULE TYPE: CDNA
35
36 US-09-014-969-1

```

Query Match	2.4%;	Score 81.8;	DB 2;	Length 2509;
Best Local Similarity	79.18;	Pred. No. 1.2e-11;		
Matches 110; Conservative	0;	Mismatches 27;	Indels 2;	Gaps 1;

QY	258	aggtgagatttctctctgtcgccagcgtgagtgcaagtgcgtgatctcgcgtcaatg	317
Db	1946	ACATGGAGTTTGCTCTTGTCGGCCAGGCTTGAGTGCATGGCCACAATCTTGGCTTACTG	2005

QY	318	caacctcaacctccaggt--cagcagatttccacactcagctcccgataagctgagat	375
Db	2006	CAACCTCTGCCTCCACAGTTCAGTGATTTCTCTGCTCCCTCAGCCCTCCCAAGTAGCTGGAT	2065

QY	376	tacagagttat	taaccac	394
Db	2066	TACAGGTA	CTCGTCA	CCAC 2084

```

RESULT 12
US-08-457-254-5/c
: Sequence 5, Application US/08457254
: Patent No. 5986049
:
: GENERAL INFORMATION:
:
: APPLICANT: Forstrom, John W
:
: APPLICANT: Lofton-Day, Catherine E
:
: APPLICANT: Lok, S1
:
: TITLE OF INVENTION: PURIFIED THROMBOPOIETIN AND METHOD OF
:
: TITLE OF INVENTION: MAKING IT
:
: NUMBER OF SEQUENCES: 24
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Zymogenetics, Inc.
:
: STREET: 1201 Eastlake Avenue East
:
: CITY: Seattle
:
: STATE: WA
:
: COUNTRY: USA
:
: ZIP: 98102
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/457,254
:
: FILING DATE:
:
: CLASSIFICATION: 514
:

```



```
ATTORNEY/AGENT INFORMATION:
: NAME: Parker, Gary E
: REGISTRATION NUMBER: 31,648
: REFERENCE/DOCKET NUMBER: 94-11C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-442-6673
: TELEFAX: 206-442-6678
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4823 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(632..644, 876..1003, 1290..1376, 3309..3476,
: LOCATION: 3713..4375)
US-08-457-254-5

Query Match          2.4%; Score 80.8; DB 2; Length 4823;
Best Local Similarity 79.0%; Pred. No. 3e-11;
Matches 109; Conservative 0; Mismatches 27; Indels 2; Gaps 1;

QY 258 aggtgagtttgcctctgtcgccagcgtgagtgagtgagtgatctcgctcactg 317
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2084 AGACGGAGTTTGTCTGTGTCAACCAGCGTGAGTGCAATTGGCATGATCTCAGCTCACTA 2025

QY 318 caacctcacctcccaagt--cagcgatttcccacttcagcctcccgataagctgagat 375
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2024 CAACCTCCACCTCTGCTTCAAGGAGATTCTCTGCTTCAGCCCTTCCAAGTAGCTGGGAT 1965

QY 376 tacagagttattaacca 393
   ||||| | ||||
Db 1964 TACAGGCATGCACCTACCA 1947

RESULT 13
US-08-484-257-20/C
: Sequence 20, Application US/08484257
: Patent No. 5989537
: GENERAL INFORMATION:
: APPLICANT: HOLLY, Richard D.
: APPLICANT: Lok, Si
: APPLICANT: Foster, Donald C.
: APPLICANT: Hagen, Frederick S.
: APPLICANT: Kaushansky, Kenneth
: APPLICANT: Kuiper, Joseph L.
: APPLICANT: Lofton-Day, Catherine E.
: APPLICANT: Oort, Pieter J.
: TITLE OF INVENTION: Methods for Stimulating Granulocyte/Macrophage Lineage Cells
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ZymoGenetics, Inc.
: STREET: 1201 Eastlake E.
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,257
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, Gary E
: REGISTRATION NUMBER: 31-648
: REFERENCE/DOCKET NUMBER: 95-09
```

```
TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-442-6673
: TELEFAX: 206-442-6678
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4823 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(632..644, 876..1003, 1290..1376, 3309..3476,
: LOCATION: 3713..4375)
US-08-484-257-20

Query Match          2.4%; Score 80.8; DB 2; Length 4823;
Best Local Similarity 79.0%; Pred. No. 3e-11;
Matches 109; Conservative 0; Mismatches 27; Indels 2; Gaps 1;

QY 258 aggtgagtttgcctctgtcgccagcgtgagtgagtgagtgatctcgctcactg 317
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2084 AGACGGAGTTTGTCTGTGTCAACCAGCGTGAGTGCAATTGGCATGATCTCAGCTCACTA 2025

QY 318 caacctcacctcccaagt--cagcgatttcccacttcagcctcccgataagctgagat 375
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2024 CAACCTCCACCTCTGCTTCAAGGAGATTCTCTGCTTCAGCCCTTCCAAGTAGCTGGGAT 1965

QY 376 tacagagttattaacca 393
   ||||| | ||||
Db 1964 TACAGGCATGCACCTACCA 1947

RESULT 14
US-08-999-927-5/C
: Sequence 5, Application US/08999927
: Patent No. 6099830
: GENERAL INFORMATION:
: APPLICANT: Kaushansky, Kenneth
: TITLE OF INVENTION: Methods of Stimulating Erythropoiesis
: TITLE OF INVENTION: Using Hematopoietic Proteins.
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ZymoGenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/999,927
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/347,748
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, Gary E
: REGISTRATION NUMBER: 31-648
: REFERENCE/DOCKET NUMBER: 94-09C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-442-6673
: TELEFAX: 206-442-6678
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4823 base pairs
: TYPE: nucleic acid
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:30:10 ; Search time 11371 Seconds
(without alignments)
2063.239 Million cell updates/sec

Title: US-09-316-048-7
Perfect score: 3348
Sequence: . 1 actctgcgcggcgctgcgg.....taatactaaaaaaaaaaaaa 3348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp14:*
121: em_estrol5:*
122: em_estrol6:*
123: em_estrol7:*
124: em_estrol8:*
125: em_estrol9:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: gb_gss7:*
159: gb_gss8:*
160: gb_gss9:*
161: em_gss5:*
162: em_gss6:*
163: em_gss7:*
164: em_gss8:*
165: em_gss9:*
166: em_gss10:*
167: em_gss11:*
168: gb_gss10:*
169: gb_gss11:*
170: em_gss12:*
171: gb_gss12:*
172: gb_gss13:*
173: gb_gss14:*
174: gb_gss15:*
175: gb_gss16:*
176: gb_gss17:*
177: gb_gss18:*
178: gb_gss19:*
179: em_gss13:*
180: gb_gss20:*
181: gb_gss21:*
182: gb_gss22:*
183: gb_gss23:*
184: gb_gss24:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1757.8	52.5	1801	146	U69197	U69197 U69197 Soar
2	701	20.9	1018	136	BE867390	BE867390 601442470
3	646.2	19.3	1036	137	BE889880	BE889880 601512132
4	621.2	18.6	724	97	AW952360	AW952360 EST364430
5	617.2	18.4	752	137	BE894295	BE894295 601437666
6	610.4	18.2	829	106	BE300155	BE300155 600944602
7	606.4	18.1	821	107	BE379687	BE379687 601159257
8	604	18.0	609	107	BE380123	BE380123 601159257
9	603.8	18.0	686	110	BE613972	BE613972 601504093
10	602.2	18.0	674	38	AV707406	AV707406 AV707406
11	599.8	17.9	665	106	BE278193	BE278193 601180017
12	594.4	17.8	616	111	BE674157	BE674157 7d76a04.x
13	588.2	17.6	594	40	AW162002	AW162002 au72c03.x
14	571.4	17.1	612	38	AV695897	AV695897 AV695897
15	565.4	16.9	749	110	BE617223	BE617223 601441906
16	555	16.6	609	136	BE810330	BE810330 MR4-PT005
17	551.6	16.5	615	134	BE082712	BE082712 RC2-BT064
18	550	16.4	550	107	BE348298	BE348298 hw16b01.x
19	548	16.4	564	40	AW163206	AW163206 au97f05.x
20	542.4	16.2	584	136	BE813737	BE813737 RC1-BN003
21	540.6	16.1	831	106	BE299828	BE299828 600944602
22	534	15.9	673	135	BE781223	BE781223 601469159
23	532.2	15.9	573	3	AA206573	AA206573 2q51f08.r
24	531.2	15.9	536	136	BE837505	BE837505 RC2-FN009
25	531	15.9	866	136	BE868328	BE868328 601443887
26	528.8	15.8	581	106	BE278342	BE278342 601179819
27	517.8	15.5	910	138	BE966164	BE966164 601660093
28	513	15.3	543	134	BE089861	BE089861 RC5-BT070
29	511.4	15.3	513	13	AA889669	AA889669 ak54a04.s
30	510.4	15.2	512	27	AI983007	AI983007 wt46g02.x
31	508.8	15.2	852	137	BE872137	BE872137 601446238
32	500.6	15.0	673	109	BE541462	BE541462 601067913
33	497.4	14.9	633	89	AW320695	AW320695 uc22b10.y
34	488.2	14.6	569	3	AA191622	AA191622 zp81g10.s
35	487	14.5	516	25	AI819766	AI819766 wj42a10.x
36	483.4	14.4	485	25	AI802592	AI802592 wf16a03.x
37	481.4	14.4	686	40	AW160378	AW160378 au72c03.y
38	478.4	14.3	480	134	BE042598	BE042598 hc26e03.x
39	475	14.2	494	89	AW351909	AW351909 RC0-CT020
40	470.4	14.1	509	142	N31181	N31181 yx64f02.r1
41	470	14.0	517	7	AA427366	AA427366 zw33c05.r
42	470	14.0	517	92	AW580332	AW580332 IL2-HT044
43	467.4	14.0	703	137	BE888420	BE888420 601514177
44	457	13.6	724	38	AV700494	AV700494 AV700494
45	455.2	13.6	561	141	H16250	H16250 yml3d10.s1

ALIGNMENTS

RESULT 1
U69197 U69197 1801 bp mRNA EST 27-OCT-1999
LOCUS U69197 Scores infant brain INIB Homo sapiens cDNA clone 22368, mRNA
DEFINITION U69197 sequence.
ACCESSION U69197
VERSION U69197.1 GI:2739420
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE 1 (bases 1 to 1801)
AUTHORS Volorio,S., Simon,G., Repetto,M., Cucciardì,M., Banfi,S., Borsani
,G., Ballabio,A. and Zollo,M.
TITLE Sequencing analysis of forty-eight human image cDNA clones similar
to drosophila mutant protein
JOURNAL DNA Seq. 9 (5-6), 307-315 (1998)
MEDLINE 99452388
COMMENT Contact: Zollo, Massimo
Telethon Institute of Genetics and Medicine
Via Olgettina 58, Milan, MI 20132, Italy
Email: zollo@tigem.it.
FEATURES
source
1.1801
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p13.1"
/clone="22368"
/clone_1db="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGGAAGATTCGGCGCGCAGGAATTTTCTTTTCTTTTCTTTT
3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 503 a 409 c 386 g 503 t
ORIGIN
Query Match 52.5%; Score 1757.8; DB 146; Length 1801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1577 gctccgaatcctgcgaagcggaagtaactgcatgatagcccgagagtggtgtatggy 1636
Db 5 GCTCCGAATCTGCCAAGGCCACGGTAACCTGCCATGATAGCCCCGAGAGTGTGTATGGG 64
QY 1637 ggcacctgcgccacagccgagacatttaagaatacatctctcagggccagtaacc 1696
Db 65 GGCACCTCGCCACAGCCGAGACATTTTAAAGATAACATCTCTTCAGGCCACGTACCC 124
QY 1697 catgacctctacgagacctctgagcaactgactatcttcagaggtccaggaatc 1756
Db 125 CATGACCTCTCAGAGACCCCTGAGCAACTGACATATCTTCCAGAGTCCAGGATTC 184
QY 1757 caggtgaatacaaaagacttccccaaaaacaacaagaagaatttgtatctcttacaat 1816
Db 185 CAGGTGATATACAAAGACTTCCCAAAAAACAACAAGAACGAATTGTATCTCTTATCAAT 244
QY 1817 tgcctctcagccacctctgcatcagccatggtatcggaagagatgtggaatcctccat 1876
Db 245 TGCTCCTCTCAGCCACCTCTGATCAGCCATGGTATCGGCAAGGATGTGAGTCTCTGCCAT 304
QY 1877 gatatgctgcgctgaacatcttaaaagtgtgtctgagtgaagccaacaagaagtaagag 1936
Db 305 GATATGGCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGGACCAACAAGTACAGAG 364
QY 1937 atgccaagaacaggaacggaacatgtctgtgtgtggaagtgctgaaccttttcggc 1996
Db 365 ATGCCAAGACAGGAACGAGCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGC 424
QY 1997 catgaaccattataaaatcccaatatatactgaanaataactgcttgaanaatt 2056
Db 425 CATGAACCATTTATAAATCCCAACATATATACTGAATAATCTGAAACTGCTTTGAATAAT 484

QY 2057 tgyaattctgatacctccagatggyccgagagacaagtggytaaggaatgtgycagca 2116
Db 485 TGGAAATTTCTGATACCTCCAGTGGGCCGAGAGACACGGTGGTAAGATGTGGCAGCA 544
QY 2117 gcaggaagacaacagaaacacaaagagcggtgtgtgcc-ggctgactgtgtggyt 2175
Db 545 GCAGGAAGACACAGAAACACAGAGAGCGGCTGTGGCGGGCTGGACTGTGCTGGGGT 604
QY 2176 ttgttgtatgycacctggtgacctggygcgtccctacgcaatagcagctgctgtggyg 2235
Db 605 TTGTTGTGATGGCCACTCGGTGACTGGCGGTCCCTACGCAATAGCAGCTGCCGTGGGG 664
QY 2236 aagaaggtgccccagccagctgttctccggygacacccagcagatccacacctggygca 2295
Db 665 AAGAAAGGCTGCCACGACGCTGTTCTCCCGGACACCAAGATCCACACCTGGGCA 724
QY 2296 cctccgtgttggtcttlttccctgtgtgaagaagaaaacygcagacccctctc 2355
Db 725 CCTCCGTGTGTGGTCTTTTTCCTGTTGTAAGAGAAACGGCACGACCCCTTCTC 784
QY 2356 aagctgctcaactcagacacatlggacaacacccctgacagccatgcccagagagagcct 2415
Db 785 AAGCTGGCTCACTCAGACACATTTGGACAACACCCTGACAGCCATGCCAGAGAGAGCCT 844
QY 2416 ttgaccgccccagagctaanaagcaccaagaanaatcaaatgcttctactcagctgtac 2475
Db 845 TTGACCGGCCCCAGAGCTAAAGACCAAGAAATCAATGCTTCTACTCAGCGTGAC 904
QY 2476 ccaacttltctagtgtgccaagccccaccacacctctctgacgtaccacacatcaccact 2535
Db 905 CCAACTTTTCTAGTGTGCCACGGCCCCACCACTCTCTGACATCCACACCATCACCAC 964
QY 2536 gcttctcttccaacagatgactgtatctttagttcattatlttctttagatgtatg 2595
Db 965 GCTTCTCTTCCAACAGTGATCTGTATTTCTTAGTTTCAATATTTTCTTTGATTGATATG 1024
QY 2596 acactataaaaatttcaatttgagaatttctcaattgtatcagttlaatatgacagtt 2655
Db 1025 ACACATATATAAAATTTTCATTTGAGATTTCTCAATTTGTATCTAGTTAAATAGCACAGTT 1084
QY 2656 tgyaaactgtcctgagactgacttcaataatctaacgcacaagaatcatatcagt 2715
Db 1085 TGGAAACTGTGCTGAGACTGACTTTATCAATAATCTAACCAGCAAAAGATCATATCCATGT 1144
QY 2716 gtatgtgttagacatttatttcaattgactaaaccagagcagtttcagtgatgcaaat 2775
Db 1145 GTATGTGTTAGACATTTTATTTCAATTGACTAACCACGAGACAGTTTCAGTGATGCAAT 1204
QY 2776 tgtgtgccctctgtgtcagctgaaaacgtccctggaacttcaaaaaaccttgaataagctc 2835
Db 1205 TGTGTGCCCTCTGTGTTCAAGTGAACAGTCTTGACTTTCAAAAACTTGAATAAGTCTC 1264
QY 2836 ccaagttgtataaaattgacaatttaggaattttaaactttagatgataatgtgtcc 2895
Db 1265 CCACAGTTGTATAAATTGGACAATTTAGGAATTTTAAACTTTAGATGATCATTTGGTTCC 1324
QY 2896 atttatttcaatttatttltgttaatgcaaacagagactttaaattgaacttgaatcct 2955
Db 1325 ATTTTATTTCAATTTTATTTTGTATGCAAAACAGAGACTTAAATGAACCTTGATCTCT 1384
QY 2956 gttttaaagattatlaaaaaaacatttgtatctatatacatagtctctttagagactagct 3015
Db 1385 GTTTTAAAGATTATTTAAAAACATGTGTATCTATACATATGGCTCTGAGGACTTAGCT 1444
QY 3016 ttcaactacactacagatatgatactcactgtagttagtccatatataaactgcagagtgatttc 3075
Db 1445 TTCACTACACTACAGATATGATGATCTTCATGTAGTCCATATATAAACCCTGCAGAGTGAATTTTC 1504
QY 3076 cagagtgctcgaactgttlaatatcacttccattagggtcgtgaaaaagaatgacctagttt 3135
Db 1505 CAGAGTGCTCGATACTGTTAATTACATCTCCATTAGGGGCTGAAAAAGAAATGACCTACGTTT 1564
QY 3136 ctgtatacagctgtgtgtcttlttagtgtgttactgtlacacagaagtgltgtcactga 3195

```
|||||
Db 1565 CTGTATACAGCTGTCTTCTTTGATGTTGTTACTGTACACAGAGTGTGTGCACTGA 1624

QY 3196 ggcctcgtgtgtgtccgtatgnaaacctgtgaccctgcgagtttaagtaactgtctcca 3255
|||||
Db 1625 GGCTCTGCGTGTGTCCGATGGAAGCCGTGTAGCCCTGCGAGTTAAGTACTGCTTCCA 1684

QY 3256 ttcatgtttacgctggaattttctcccatgtgaatgttaagtaaaacttaagtgtt 3315
|||||
Db 1685 TTCAATTGTTACCGTCGAATTTTCTCCCATGGAATGTAAGTAAACTTAAGTGTGTGT 1744

QY 3316 catcaataatgttaataactaaaaa 3348
|||||
Db 1745 CATCAATAATGTTAATAAAAAA 1777

RESULT 2
LOCUS BE867390 1018 bp mRNA EST 27-SEP-2000
DEFINITION 601442470F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846689 5',
mRNA sequence.
ACCESSION BE867390 GI:10316166
VERSION BE867390.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1018)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM546 row: 1 column: 18
High quality sequence stop: 654.
FEATURES
source
1. 1018
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846689"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: Notti;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 295 a 267 c 262 g 194 t
ORIGIN
Query Match 20.9%; Score 701; DB 136; Length 1018;
Best Local Similarity 90.0%; Pred. No. 6.8e-147;
Matches 809; Conservative 0; Mismatches 80; Indels 10; Gaps 5;

QY 1122 tggccagatccagcagcaaaaaagagaagagccagagatcacgctcctcacagagc 1181
|||||
Db 1 TGGCCAGATCCAGCAGCAAAAGAGAGAGCAGAGTACACGCTCCTCACAGAGC 60

QY 1182 gaggcctccgcgcgcagggaggtttgtgatgcaggtgaaggttggaaccacactgcag 1241
|||||
Db 61 GAGGCCCTCCCGCCGCAAGGAGTTGTGTATGCAAGGTGAAGGTTGGAAACACACACTGCAG 120

QY 1242 aaggaacggygcaccaacaagaaggtgccaagcgcaatgcagccgagaacatgtgtgaga 1301
```

```
|||||
Db 121 AAGGAACGGGCACCAACAAGAAGGTGGCCAGCGCAATGCAGCCGAGAACATGTGAGAGA 180

QY 1302 tccttggttcaaaagtcctgcagcgcgcagcccaaacccgcactcaagtcagagaga 1361
|||||
Db 181 TCCTTGTTTCAAAAGTCCCGCAGGCGCAGCCACCAAAACCGACTCAAGTCAGAGAGA 240

QY 1362 agacaccataaagaaccagggatgtgaaagaaagtaacctttttgaacctgctctg 1421
|||||
Db 241 AGACACCCATTAAGAACACCAGGGGATGGAAGAAAGTAACCTTTTGAACCTGGCTCTG 300

QY 1422 gggatgaaaatgaggactagtaataagagagatggttcagagatgccttatcgaatc 1481
|||||
Db 301 GGGATGAAATGGGACTAGTATAAAGAGGATGAGTTCAAGATGCCCTTATCTAAGTCATC 360

QY 1482 agcagctgcctgctggaattctcccatgtgcccgaaggtcgcccaggtctgagagta 1541
|||||
Db 361 AGCAGCTGCCTGTGGAATTTCTCCCATGTGTCCCGAGGTGCGCCAGGCTGTAGAGTTA 420

QY 1542 gtcaagacatcacaccaaaagatttaccagggcagctccgaatcctcgcagagccagcg 1601
|||||
Db 421 GTCAAGACATCACACCAAAAGATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGG 480

QY 1602 taactgccatgatagcccgagaggttgtgtatgtggggcaccctgcgccacagccgagaca 1661
|||||
Db 481 TAACTGCCATGATAGCCCGAGAGTGTGTCTATGGGGGCACCTCGCCACAGCCGAGACCA 540

QY 1662 tttaaagaataacatctcttcagggccagtcacccatgagctctcagagaccctctg 1721
|||||
Db 541 TTTTAAAGAATAACATCTCTTCAGGCCACGTAACCCATGAGACCTCTTCACGAGACCCTCTG 600

QY 1722 agcaactgactatcttccagagtcagagattccaggttgaatacaaaagactcccca 1781
|||||
Db 601 AGCAACTGGACTATCTTCCAGAGTCCAGGATTCAGGTTGAATACAAAGACTTCCCCA 660

QY 1782 aaacaacaagaagaattgtatct--ctatcaattgtctcctcagccactctga 1838
|||||
Db 661 AAAATCACCAAGAAGCAATTCGATATCTCTTATCAATGAGCTCTCTCAGCCACTGTGAT 720

QY 1839 tcagccatggtatcggaagatgtgtgagtcctgcca--tgatatggtgcgtgaaca 1895
|||||
Db 721 CCAACATGTTATCGGACAGGTGTGTGGGAATCTCGCAATGATGATGGGTGGCTGAGACA 780

QY 1896 tcttaag-ttgcctgtcgtgagttgaccacaacaa-gtacagagatgccaagaagagaa 1953
|||||
Db 781 TCTTTAAGATGGCTGCGGGAAGTGGACCCACAAAGGTCAAAAAATTTCCAGAAACGGGAA 840

QY 1954 cggaccaatgtctgtgtgtgg--agtgctgaaaccttcttgccatgaaccattata 2010
|||||
Db 841 GGGACCCAGTTGGTGTGTTGGCAGACAGGTCTCGAACCAATTTTGGGCGCTGAACCATTA 899

RESULT 3
LOCUS BE889880 1036 bp mRNA EST 29-SEP-2000
DEFINITION 601512132F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3913590 5',
mRNA sequence.
ACCESSION BE889880
VERSION BE889880.1 GI:10347645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1036)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>

Plate: LNCM720 row: P column: 07
High quality sequence stop: 706.

FEATURES

location/Qualifiers
1. 1036

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3913590"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 296 a 230 c 220 g 290 t
ORIGIN

Query Match 19.3%; Score 646.2; DB 137; Length 1036;
Best Local Similarity 91.5%; Pred. No. 1.4e-134;
Matches 740; Conservative 0; Mismatches 63; Indels 6; Gaps 5;

QY 2421 cggccccagagctaaagcaccagagaaatcaaatgcttcttactcagcgtgacccaac 2480
Db 1 CGGCCCCAGAGCTAAAGCACCAGAGAAATCAATGCTTCTACTCAGCGTGACCCAAC 60
QY 2481 ttctctagtgtgccacgycgccccaccctctctgagtaaccacatcaccactgctt 2540
Db 61 TTTTCTAGTGTGCCACGGCCCCACCACCTCTGCAGTACCACCATCACCACCTGCTT 120
QY 2541 ctcttccaacagtgatctgtatcttctgttcttatttcttcttcttcttcttcttct 2600
Db 121 CTCTTCCAACAGATGATCTGTATTTCTAGTTTCATTTCTTTGATTTGATATGACACT 180
QY 2601 atataaaatttcatttgagaatttcctaattgtatctagttaaatagcacagtttgaa 2660
Db 181 ATATAAATTTTCATTTGAGAATTTCTCAATTTGTATCTAGTTAAATAGCACAGTTTGAA 240
QY 2661 actgtctgagactgacttatacaataatcctaaccgacaagaatcataatcattgtatg 2720
Db 241 ACTTGTCTGAGACTGACTTTATCAATATCTTAACCGACAAGATCATATCCATGTGTATG 300
QY 2721 tggtagacattttatttattcattgactaaccacagagcagtttcaagtatgcaaatgtgt 2780
Db 301 TGGTTAGACATTTTATTTTCATTGACTTAACCCAGGACAGTTTCAAGTATGCAAAATGTGT 360
QY 2781 gccctctgttcagctgaaacagtcctctgactttcaaaaacctgaaataagctccacaca 2840
Db 361 GCCCTCTGTTGAGCTGAACAGTCCCTGACTTTCAAAAACCTGAATTAAGTCTCCACACA 420
QY 2841 gtgtataaatggaacaatttaggaattttaaactttagatgatcatttggttccattt 2900
Db 421 GTGTATAAATGGAACAATTAGGAATTTTAAACCTTAGATGATCATTTGGTTCCATTTT 480
QY 2901 tattcatttttattttgttaatgcaacagacttaaatgaactttgatctctgtttt 2960
Db 481 TATTTCATTTTATTTTGTGTAATGCAACAGGACTTAATGAGACTTTGATCTCTGTTTT 540
QY 2961 aaagattataaaaaaacattgtgtatctctatatacatatgctccttgagagacttagcttccac 3020
Db 541 AAAGATTATTAATAAACATTTGTGTATCTATACATATGGCTCTTGAGGACTTAGCTTTCA- 599
QY 3021 taactacagatagatctcc-atgtagtcataataaacctgcagagtg--atttcca 3077
Db 600 TACACTACAGATATGATCTCCCAATGTAGTCCATATTAACCTGCAGAGTGGATTTTCCAG 659
QY 3078 gagtgcgcgact-gttaattacatctcatttagggtgtaaaagaatgacctagcttcc 3136
Db 660 AAGTGTGATGATGCTTCATTTACTCTCCATTTAGGGTGAAAAAGAAATGGACCTAGTTCTG 719

QY 3137 tgtatacagctgtgtgtcttlttgatgttgtgttactgttacacagaagtgtg-tgcactga 3195
Db 720 TATAAAGTGTGTGTGCTTTCAGAGTTGCGGGTTACTGTGTAACACAGACTGTGTCCCTGA 779

QY 3196 ggctctgcgtgtgtccglatggaacc 3224

Db 780 GGGTCTGATGTGGCCCGCTGTATACC 808

RESULT 4
AW952360 724 bp mRNA EST 01-JUN-2000
LOCUS
DEFINITION EST364430 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.
ACCESSION AW952360
VERSION AW952360.1 GI:8142042
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 724)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 38
Seq primer: Reverse.

FEATURES
location/Qualifiers
1. 724

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGB"
/note="Vector: pBluescriptSkm"

BASE COUNT 198 a 179 c 189 g 158 t
ORIGIN

Query Match 18.6%; Score 621.2; DB 97; Length 724;
Best Local Similarity 95.1%; Pred. No. 5.3e-129;
Matches 685; Conservative 0; Mismatches 28; Indels 7; Gaps 4;

QY 1581 cgaatcctgcgaagcgcaagtaactgcatgatagccgagagttgtgtatggygca 1640
Db 1 CGAATCCTGCGCAAGGCCAGGTAAGTCCATGATAGCCGAGAGTGTGTATGGGGCA 60
QY 1641 cctcgccacagccgagacattttaagaataacatctctcagggccagtaaccatg 1700
Db 61 CCTCGCCACAGCCGAGACCATTTTAAAGAAATACATCTCTTCAGGCCACGTACCCCATG 120
QY 1701 gaccttcacgagaccctctgagcaactgactatcttccagagtcaggagattccag 1760
Db 121 GACCTCTCAGAGACCCCTCTGAGCAACTGACATATCTTTCAGAGATCCAGGATTCAGG 180
QY 1761 ttgaatacaaaagacttccccaaaaaacaacaagaacgaatttgtatctctatcaattgct 1820
Db 181 TTGAATACAAAGACTTCCCAAAAACAACAAGAAGCAATTGTATCTCTTATCAATTGCT 240
QY 1821 cctctcagccacctctgatacagcatggtatgycgcaagagtgtgagtcctgcatgata 1880
Db 241 CTCTCAGCCACCTCTGTATCAGCCATGTATGGCAAGATGTGAGATCTGTCATGATA 300
QY 1881 tggctgcgtgaacatcttaagttgtgtctgtgagttgagcaacaagttacagatgc 1940
Db 301 TGGCTGCGCTGAACATCTTAAAGTTGCTGTGTGAGTTGGACCAACAAGTACAGAGATGC 360

Tissue Procurement: ATCC
CDNA Library Preparation by: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM54 row: d column: 02
High quality sequence start: 41
High quality sequence stop: 799.

FEATURES

Source

1. .829
Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2960593"

/clone_lib="NIH_MGC_17"

/tissue_type="rhabdomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;

Site_2: XhoI; CDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-CDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

190 a 200 c 226 g 213 t

ORIGIN

Query Match

Best Local Similarity 18.2%; Score 610.4; DB 106; Length 829;

Matches 622; Conservativity 0; Mismatches 1; Indels 1; Gaps 1;

QY 378 cagagtttataaccacttaacctctcagaactgaaacacacatgttctcgtgaac 437
Db 649 CAGAGTTTATTAAACCACTTAACCTCTCAGAACGACAAAGACAACATGTCTCTGG-AC 591
QY 438 gccctcttttaaaaaaagacataaccctactgtagaactaatgtcactgtgcatga 497
Db 590 GCCCTCTTTTAAAAAAGAACATTAACCCCTACTGTAGAATAATGCACTGTGCATGA 531
QY 498 aacttgaaaaaaaccaatgtataagcctgttgacccttactctcgtatgcagtcacact 557
Db 530 AACTTGAAAAAACCAATGTATTAAGCCTGTGACCCCTTACTCTCGGATGCAGTCACCT 471
QY 558 ataactacaacatgagagagtgctatccccgaggtactttaccattccagttc 617
Db 470 ATAACTACAACATGAGAGAGGTGCTTATCCCCGAGGTACTTTTACCACATTTCCAGTTC 411
QY 618 cacccttacttatacaagtgaactttctgtggagagacagcaatttaatggcaagaa 677
Db 410 CACCTTTACTTTATCAAGTGAACCTTCTGTGGAGAGACAGCAATTTAATGCAAAAGGAA 351
QY 678 agacaagacagctgcgaaacaacgatgtctgtccaaagcgttgagatcctgcagaatg 737
Db 350 AGACAAGACAGAGCTGCGAAACACGATGCTGTGCGCAAAAGCGTTGAGGATCTGCAAGATG 291
QY 738 agccccctgcagagagctgaggtgaatgtaagagaalccgaagaagaatctcaata 797
Db 290 AGCCCCCTGCAGAGAGGCTGAGGTGAATGGAAGAAATCCGAAGAAGAAATCTCATATA 231
QY 798 aatctgaataagtcaagtgtttgagatgtgacctaaacggaactgtcctgtgaattcg 857
Db 230 AATCTGAATAATAGTCAAGTGTGAGATGTGACCTTAACGGAACCTGCTGTGAATTTGCG 171
QY 858 aggtgccccggagagagtgccccacccacatgaagaactttgtgaccaaggtttcggttg 917
Db 170 AGGTGCCCCGGAGAGTGGCCCCACCCACATGAGAAGCTTTGTGACCAAGGTTTCGGTTG 111
QY 918 gggagtttggtggggaagtgaagggaagaagatttcaagaagaatgcgcga 977
Db 110 GGGAGTTTGTGGGGGAAGGTGAAGGGAAGCAAGAGATTTCAAAGAAATAATGCCGCCCA 51

QY 978 tagctgttcttgaggagctgaaga 1001
Db 50 TAGCTGTTCTTGAGAGAGCTGAAGA 27

RESULT 7

BE379687/c

LOCUS

DEFINITION

BE379687 821 bp mRNA EST 21-JUL-2000
601159257T1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510806 3',

mRNA sequence.

BE379687

BE379687.1 GI:9325052

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM196 row: a column: 15
High quality sequence start: 19
High quality sequence stop: 755.

location/Qualifiers

1. .821

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3510806"

/clone_lib="NIH_MGC_53"

/tissue_type="carcinoma, cell line"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgccctggcc); Site_2: SfiI (ggccatattggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGGCGGAGGCGGCGGACATG-dT(30)BN-3'
sequence: 5'-CAGGCGCATTTATGGCC-3' and 3' adaptor
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT

291 a 154 c 150 g 226 t

ORIGIN

QY 2651 cagttggaactgtctgagactgacttatac-aataatctaaccgacaagatcatat 2709
Db 820 CACAGTTGGAAGAACTGCTGAGACGGAGCTGTATCGAAAAATCTAACCGACGAAGATCATAT 761
QY 2710 ccatgtatgtgttagacattttat-ttcattgactaaccagagacgtttcagltga 2768
Db 760 CAATGTGTATGTGTTAGACATTTTATGTTCATTGACTAACCAGAGACAGTTTCAGTGA 701
QY 2769 tgcgaattgtgtccctctgtgtcagctgaa-acagtcctgagac--ttcaaaaaccttg 2825
Db 700 TGCAAATTGTGTGCCCTCTGCTTCAGCTGAAGACAGACGTCCTGAGCGTTTCAAAAACCTTTG 641
QY 2826 aataagctctccacagttgtataaattgagacaatttaggaattttaactttagatgac 2885

Query Match 18.1%; Score 606.4; DB 107; Length 821;

Best Local Similarity 96.8%; Pred. No. 1.1e-125;

Matches 672; Conservativity 0; Mismatches 16; Indels 6; Gaps 5;

|||||
Db 640 AATAAGTCTCCACAGTGTATATAATGGACAATTAGGAATTTAACTTAGATGATC 581
QY 2886 atttggtccatt-ttattcatttttatttggtaaatgcaaacaggacttaaatgaa 2944
Db 580 ATTTGGTTCATGTGTTATTTATTTTGTATGTAATGCAACAGGACTTAATGAA 521
QY 2945 cttgatctctgtttaaagattatataaaacattgtgtatcatacatatgctcttg 3004
Db 520 CTTTGATCTCTGTTTAAAGATTATTAACATTGTGTATCTATACATATAGGCTCTTG 461
QY 3005 aggaactagcttcaactaactaactaactaactatccatgtagtcacataaactgca 3064
Db 460 AGGACTTAGCTTCACTACACTACAGGATATGATCTCCATGTAGTCCATATAAACCCTGCA 401
QY 3065 gaagtatttccagagtgctgcatactgttaatacatctccattaggcgctgaagaat 3124
Db 400 GAGTGATTTCCAGAGTGCCTCGATACGTGTAATTACATCTCCATTAGGGCTGAAAGAAT 341
QY 3125 gaactacgttctgtatatacagctgtgtgctttgtatgtgtgttactgttacacagaat 3184
Db 340 GACCTACGTTCTGTATACAGCTGTGTGCTTTTGATGTGTGTACTGTACACAGAGAT 281
QY 3185 gtgtgcaactgaagctctgcgtgtgtgcgtatggaacacctgtagccctgcgagttaag 3244
Db 280 GTGTGCACTGAGGCTCTGCGTGTGCTGTATGGAAGCCCTGTAGCCCTCGAGTTAAG 221
QY 3245 tactgtccatcatctgtttacgctggaatttctcccatggaatgttaagtaaac 3304
Db 220 TACTGCTTCCATCATGTTGTTACGCTGGAATTTTCTCCCATGGAATGTAAGTAAACT 161
QY 3305 taagtgttgcacataaataatgtaataactaaa 3338
Db 160 TAAAGTTTGTATCAATAAATGTAATACTAAA 127

RESULT 8
BE380123 609 bp mRNA EST 21-JUL-2000
LOCUS 601159257F2 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510806 5',
DEFINITION mRNA sequence.
ACCESSION BE380123
VERSION BE380123.1 GI:9325488
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 609)
REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LCM196 row: a column: 15
High quality sequence stop: 606.

FEATURES
source
1. 609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3510806"
/clone_lib="NIH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site_1: SfilI (ggccgcctcgcc); Site_2: SfilI (ggccattatggcc

); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATATGCCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGGGCGAGCGGCCGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 178 a 125 c 96 g 210 t
ORIGIN

Query Match 18.0%; Score 604; DB 107; Length 609;
Best Local Similarity 100.0%; Pred. No. 3.8e-125;
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2395 agccatgccagagagagagcccttggaccgccccagagctaaagcaccagagaataatcaa 2454
Db 1 AGCCATGCCAGAGAGAGGCGCTTGCACCGGCCCCAGAGCTAAAGACACAGAGAAAATCAA 60
QY 2455 atgcttcactaagcgtgacccaactttcttagtgtgtgccagggccccaccactctgc 2514
Db 61 ATGCTTCTACTACAGCGGTGACCCAACTTTTCTAGTGTGCCACGGCCCCACACACCTCTGC 120
QY 2515 agtaaccacacacacacacactgtcttcttccaacagtgatctgtatctttagttcat 2574
Db 121 AGTACCACACACATCACCACTGCTTCTCTTCCAAACAGTAGATCTGATTTCTTAGTTTCA 180
QY 2575 tatttcttltgatatgacactatataaatttcatattgagaatttctcaattgt 2634
Db 181 TATTTCTTTTGAATGATATGACACTATATAAATTTTCATTTGAGAAATTTCTCAATTGT 240
QY 2635 atctaagtaaatgacacagtttggaaactgtctgagactgacttatacaataatcacc 2694
Db 241 ATCTAGTTAAATAGCACAGTTTGGAACCTGTCTGAGACTGACTTTATCAATAATCTAAC 300
QY 2695 cgacaaagatcatatccatgtgtatgtgtagacattttattcatctgactaaccag 2754
Db 301 CGACAAAGATCATATCCATGTGTATGTGTTAGACATTTTATTTCAATTGACTAACCCAG 360
QY 2755 gacagtttcaagtatgcaaatgtgtgcccctctgtgttcagctgaaacagtcctgacttt 2814
Db 361 GACAGTTTCAGTATGCAAAATGTGTGCCCTCTGTTACAGCTGAACAGTCTCGACTTT 420
QY 2815 caaaaacctgaataagtcctccacagttgtataaattggaacatttaggaattttaa 2874
Db 421 CAAAAACCTTGAAATAGTCTCCACAGTGTATTAATTTAGCAATTTTAAAC 480
QY 2875 tttagatgacattgtgtccatttttattcatcttatttggtaaatgcaacagga 2934
Db 481 TTTAGATGATCATTTGTTCCATTTTATTTTATTTTATTTTGTATGCAAAACAGGA 540
QY 2935 cttaaatgaacttgatctctgttttaagattataaaacacattgtgtatcatacat 2994
Db 541 CTTAATGAACCTTGATCTCTGTTTAAAGATTATTAAGAAACATTTGTATCTATACAT 600
QY 2995 atgg 2998
Db 601 ATGG 604

RESULT 9
BE613972 686 bp mRNA EST 24-AUG-2000.
LOCUS 601504093F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5',
DEFINITION mRNA sequence.
ACCESSION BE613972
VERSION BE613972.1 GI:9895569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 686)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM700 row: j column: 24
High quality sequence stop: 675.

FEATURES

Source
1. 686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3905783"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 226 a 156 c 196 g 108 t
ORIGIN

Query Match 18.0%; Score 603.8; DB 110; Length 686;
Best Local Similarity 97.1%; Pred. No. 4.3e-125;
Matches 669; Conservative 0; Mismatches 12; Indels 8; Gaps 5;

QY 696 aacacgatgtctgcacaaagcgttgagatcctgcagaatgagccctccagaagaagc 755
Db 1 AACACGATGCTGCTGCCAAAGCGTTGAGATCCTGCAGATGAGCCCTGCCAGAGAGGC 60
QY 756 tggaggtgaatggaagagaatcgaagaagaataatcgaataatcgaataatgaatcaag 815
Db 61 TGGAGGTGAATGGAAGAGAAATCCGAAGAAGAAATCTCAATAAATCTGAATAAATGAATCAAG 120
QY 816 tgtttgagattgcacttaacggaacttgctgctgtgaatttcgagtggtgcccggagagtg 875
Db 121 TGTTTGAGATTGCACCTTAACGGAAGTCCCTGTGTAATTCGAGGTGCCCGGAGAGAGTG 180
QY 876 gcccaaccacatgaagaactttgtgaccaaggttcggttggtggagtttggggggaag 935
Db 181 GCCCACCCACATGAAGAAC-TTGTGACCACCAAGGTTTCGGT--GGGAGTTGTGGGGGAAG 237
QY 936 gtgaagggaagaagaagaattcaagaagaataatgcgcagcatagctgttcttgaggagc 995
Db 238 GTGAAGGGAAAGCAAGAGATTTCAAAGAAAATGCCCATAGCTGTCTTGAGGAGGC 297
QY 996 tgaagaagttacgcgccttgctgagttgaagcagtaaaagcctagaaatcaaaaagaanaa 1055
Db 298 TGAAGAAGTTACCGCCCTGCTGAGTTGAACGAGTAAAGCCTAGAAATCAAAAAGAAAA 357
QY 1056 caaaaccatagtcgaagcacagacaagccagaaatgcccagggatcaatccgatga 1115
Db 358 CAAAAACCATAGTCAAGCCACAGACAAGCCAGAAATATGGCCA--GGGATCAATCCGATTA 416
QY 1116 gccgactgccccagatccagcaggaacaaaagagaaggaagccagagatcacgctctca 1175
Db 417 GCCGACTGGCCACAGATCCAGCAGGCAAAAAGAGAGAGAGCCAGAGTACACGCTCTCA 476
QY 1176 cagagcgaagcctccgcgcgcgcagggagtttgtatgcaggtgaaggttggaaccaca 1235
Db 477 CAGAGCGAGGCGCTCCGCCGCCAGGAGTTGTGATGCAAGGTGAAGGTTGGAACACACA 536
QY 1236 ctgcagaagaagcggcacaacaagaaggtgccaagcgcaatgcagccgaagaacatgc 1295

|||||
Db 537 CTGCAGAGGAACGGGGCACCAACAAGAAGGTGGCCAAAGCAATGACGCCGAGACATGC 596
QY 1296 tggagatccttgg-ttcaaatgcccgagcggcagcccaaccacccgactcaagtca 1354
Db 597 TGGAGATCCTTGGTTTTCAAAGTCCCGCAGCGCAGCACCAAAA--CCGACTCAAGTCA 653
QY 1355 gaggagaagacaccataaagaacagg 1383
Db 654 GAGGAGAGACACCCATTAAGAACAAGG 682

RESULT 10

AV707406 674 bp mRNA EST 09-OCT-2000
LOCUS AV707406 ADB Homo sapiens CDNA clone ADBCMG03 5', mRNA sequence.
DEFINITION AV707406
ACCESSION AV707406
VERSION AV707406.1 GI:10724671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 674)
AUTHORS Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
, H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
, G., Hu,R., Chen,J., Chen,Z. and Han,Z.
Homo sapiens CDNA ADB clones
Unpublished (2000)

TITLE Contact: Zeguang Han
JOURNAL Chinese National Human Genome Center at Shanghai
COMMENT 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
Source
1. 674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADBCM03"
/clone_lib="ADB"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 197 a 148 c 113 g 215 t 1 others
ORIGIN

Query Match 18.0%; Score 602.2; DB 38; Length 674;
Best Local Similarity 98.4%; Pred. No. 9.7e-125;
Matches 618; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 2352 tctcaagctggctcactcagacacattgggacaacacctggacagccatgcccagagag 2411
Db 1 TCTCAAGCTGGCTCACTCAGACACATTGGGACAAACCTGGACAGCCATGCCAGAGAGAG 60
QY 2412 gcccttgaccggccccagagctaaagaacaccagagaagaatcaaatgtctcactaacgc 2471
Db 61 GCCTTGACCGGGCCAGAGCTAAAGCACACAGAGAAATCAATGCTTCTTACTACGCG 120
QY 2472 tgaccaactttctagtgtgccaagccccaccacctctgcagtlaccacaccatcac 2531
Db 121 TGACCAACTTTTCTAGTGTGCCAGGCGCCACACACCTCTGCAAGTACACACATCAC 180
QY 2532 cactgttctctctccaacagtgatctgtatcttagttcatatattctttagatga 2591
Db 181 CACTGCTTCTCTTCCAAACAGTGAATCTGTATTCTTAGTTCAATTATTTCTTTGATTGA 240

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 469.

FEATURES

source

1. 616

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3278862"

/clone_lib="NCI_CGAP_Lu24"

/tissue_type="carcinoid"

/lab_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized

library NCI_CGAP_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was

used as tracer in a subtractive hybridization reaction.

The driver was PCR-amplified cDNAs from a pool of 5,000

clones made from the same library (cloneids

1414920-1417991 and 1520904-1522439). Subtraction by Bento

Soares and M. Fatima Bonaldo. "

BASE COUNT 179 a 153 c 139 g 141 t 4 others

ORIGIN

Query Match 17.8%; Score 594.4; DB 111; Length 616;

Best Local Similarity 98.4%; Pred. No. 5.4e-123;

Matches 598; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1361 aagacaccataaagaaccaggagatggaagaagaaccttttgaactgctct 1420

Db 8 AAGACACCCATTAAGAAACCAGGGGATGGAAGAAAGTAACCTTTTGAACCTGCTCT 67

OY 1421 ggggatgaaaatgggactagtaaaaggatgagtcagatgacctatcgaatc 1480

Db 68 GGGGATCAAAATGGGACTAGTAATAAGAGGATGAGTTCAGATGCCCTTATCTAAGTCAT 127

OY 1481 cagcagctgcctgctggaattctcccatggtgcccaggatcgccagcgtgtagagt 1540

Db 128 CAGCAGCTGCCTGCTGGAATCTTCCCATGCTGCCGAGTCCGCCAGGCTGAGAGTT 187

OY 1541 agtcaaggacatcacaccaagaatttaccaggcagctccgaatcctgccaagccagc 1600

Db 188 AGTCAAGGACATCACACCAAGAATTTTACCAGGGCAGCTCCGAATCTTGCCAAGGCCACG 247

OY 1601 gtaactgccatgatagcccagagatgtgtatggygggcaacctgcgccacagccagacc 1660

Db 248 GTAACCTGCCATGATAGCCCCGAGAGTTGTGTATGGGGGCCACCTCGCCACAGCCGAGACC 307

OY 1661 atttaagaataaacaatcctctcaagccacgtaccccatggaacctctcacgagacctct 1720

Db 308 ATTTTAAGAATAACATCTCTTCAAGGCCACGTCACCCCATGACCTCTCAGAGACCTCTCT 367

OY 1721 gagcaactgactatcctccagagtcacaggatccaggttgaaatacaagaactccccc 1780

Db 368 GAGCAACTGCACTATCTTCCAGAGTCCAGGATTCAGGTTGAATACAAAGACTTCCCC 427

OY 1781 aaaaacaagaagaagaatttgtatcttatacaattgtctcctcagccacctgtatc 1840

Db 428 AAAAACAACAAGAAGCAATTTGTATCTCTTATCAATTCCTCTCANCCACCTCTGATC 487

OY 1841 agccatgtatcgcgaagatgtgagctcctgcctcatgatatgctgcgtgaacatctta 1900

Db 488 AGCCATGGTATCAGCAAGGATGTGAGTCTTCCATGATATGCGCTGGAACATCTTA 547

OY 1901 aagtcgtctctgagttgacccaacaagtlacagagatgccaagaagaagaaagaccca 1960

Db 548 GAGTTGCTGTCTGATGTGACCAACAANTACAGAGATGCCAAGAACANGANACAGACCA 607

OY 1961 atgtctgt 1968

Db 608 ATGTCTGT 615

RESULT 13

AW162002/c

LOCUS

DEFINITION

AW162002 594 bp mRNA EST 09-NOV-1999

au72c03.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone

IMAGE:2781796 3' similar to TR:095793 095793 STAUEN PROTEIN. ;

mRNA sequence.

AW162002 GI:6301035

EST.

human.

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 594)

Hillier,L., Allen,M., Bowles,L., Dubugue,T., Geisel,G., Jost,S.,

Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maira,M., Martin

,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,Y., Wylie,T., Waterston,R. and Wilson,R.

Unpublished (1997)

Other ESTs: au72c03.y1

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 464.

FEATURES

source

1. .594

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2781796"

/clone_lib="Schneider fetal brain 00004"

/sex="male"

/tissue_type="frontal lobe"

/dev_stage="5 months post-conception"

/lab_host="DH10B"

/note="Organ: brain; Vector: Bluescript SK (Stratagene);

Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was

prepared from human fetal brain tissue. 5' and 3'

adaptors were used in cloning as follows: 5' adaptor

sequence:

5'-GAGAGAGAGAAGAGCTCAAGATCCTTAATTAATTAATCCCCCCCCCC-3'

and 3' adaptor sequence:

5'-GAGAGAGAGAGCTCGAGTTTCTTTTCTTTTCTTTT-3'. The library was

size-selected for >0.5 kb inserts and has an average

insert size estimated at 1.2 kb. This library was

constructed using the CAP-trapper method for full-length

enrichment and has not undergone amplification. Library

was constructed by Dr. Claudio Schneider (LNCIB-Area

Science Park, Trieste, Italy). "

BASE COUNT 130 a 152 c 123 g 188 t 1 others

ORIGIN

Query Match

Best Local Similarity 99.3%; Pred. No. 1.3e-121;

Matches 590; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 455 gaaagcataaccctactgtagaactaaatgcaactgtgcatggaacttgaaaccca 514

Db 594 GAAAGCATAAACCCCTACTGTAGAAGTAATGTCACTGTGCATGAACCTTGAAAAAACCCA 535

QY 515 atgtataagcctgttgacccttactctcgtatgcagtcaccctataactacaacatgaga 574
|||||
Db 534 ATGTATAAGCCTGTTGACCCTTACTCTCGATGTCAGTCCACCCTATACTACAACATGAGA 475
QY 575 ggaagtgtatccccgaggtactttaccattccagttccaccttacttataca 634
|||||
Db 474 GGAGGTGCTTATCCCCGAGGTACTTTTACCATTCCAGTTCCACCTTACTTATCA 415
QY 635 gtggaacttctgttgaggagacagcaatttaatgycaaaggaagacagacagctgcg 694
|||||
Db 414 GTGGAACCTTCTGTGGGAGACAGCAATTTATGGCAAAGGAAGACAGACAGGCTGCG 355
QY 695 aaacacgatgtctgtcccaaaagcgttgagatcctgcagaatgagccccctgcagagag 754
|||||
Db 354 AAACACGATGCTGCTGCCAAGCGTTGAGGATCTTCAGAAATGAGCCCCCTGCCAGAGAG 295
QY 755 ctggaagtgtaatggaagagaatccgaagaagaaaaatcctaataatctgaataaagtcaa 814
|||||
Db 294 CTGGAGGTGAATGGAAGAGAAATCCGAAGAAGAAATCTCAATTAATCTGAAATTAAGTCAA 235
QY 815 gtgtttgagattgcacttaaacggaactgtcctgtgaatttcgaggtgccccggagagt 874
|||||
Db 234 GTGTTTGAGATGCACTTAACCGGAACCTGCTGTAATTTGAGGTGGCCCCGGAGAGT 175
QY 875 ggccccccccacatgaagaactttgtgaccaaggttcggttgaggagttgttggggaa 934
|||||
Db 174 GGCCCAACCCACATGAGAAGACTTGTGACCAAGGTTTCGGTTGGGAGTTGTGGGGGAA 115
QY 935 ggtgaagggaagaaagcaagaagattccaagaaaaaaatgcgcgcatagtctgttcttgaggag 994
|||||
Db 114 GGTGAAGGGAAGAAAGCAAGAAGATTTCAAAGAAAAATGCCGCATAGCTGTCTTGAGGAG 55
QY 995 ctgaagaagttaccgccccctgcctgcagttgaaagagtaaaagcctagaatacaaa 1048
|||||
Db 54 CTGAAGAAGTTACC GCCCTGCTGCAGTTGAACGAGTAAGCCTGAGAAAAAAA 1

RESULT 14
LOCUS AV695897 612 bp mRNA EST 25-SEP-2000
DEFINITION AV695897 GKC Homo sapiens cDNA clone GKCEE08 5', mRNA sequence.
ACCESSION AV695897
VERSION AV695897.1 GI:10297760
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 612)
AUTHORS Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
, Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
TITLE
JOURNAL
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1. 612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCEE08"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:

BASE COUNT 149 a 164 c 125 g 174 t
XhoI"
ORIGIN

Query Match 17.1%; Score 571.4; DB 38; Length 612;
Best Local Similarity 98.4%; Pred. No. 7.9e-118;
Matches 611; Conservative 0; Mismatches 1; Indels 9; Gaps 3;

QY 2194 ggtgacctgycgtccctacgcgaatagcagctgcctgttgggaagaagggctgccagcc 2253
|||||
Db 1 GGTGACCTGGCGGTCCTTACGCAATAGCAGCTGCTGTGGGAAGAGGCGCTGCCAGGCC 60
QY 2254 agctggtctcccgagacaccagcagatccacaccctgggacctccgttgtgtctt 2313
|||||
Db 61 AGCTGTTCTCCCGGACACACAGATCCACACCTGGGCACCTCCGTGTTGTTT 120
QY 2314 ttttccccctgttgtaaagaagaacgcgcagaccccccttctcaagctggtcactcagac 2373
|||||
Db 121 TTTTCCCTGTGTGAAGAAGAAGAACGGCAGACCCCTTCTCAAGCTGCTCAGTCAAGAC 180
QY 2374 acattggacaacacctgagacagccatgcccagagagagccttggaccggccccagagct 2433
|||||
Db 181 ACATTGGACAAACCTTGAGACAGCCATGCCAGAGAGAGGCTTTGACCGGCCCCAGAGAGCT 240
QY 2434 aaaaagcaccagagaagaatacaaatgcttcctactcagcgtgacccaacttcttagtgc 2493
|||||
Db 241 AAAAGCACAGAGAAATCAAAATGCTTCTTACTCAGCGTGACCCCAACTTTTCTAGTGTGC 300
QY 2494 caagccccaccacctctcgcagtaaccacacatcaaccaactgcttctctccaacagt 2553
|||||
Db 301 CACGGCCCCACACCTCCTGAGTACCCACACCATCACCACTGCTTCTTCCAACAGT 360
QY 2554 gatctgtatcttagtttcatlatacttcttcttgatgtatgacactatataaaatttc 2613
|||||
Db 361 GATCTGTATTCTTAGTTTCAATTATTCTTTTGATGTATGACACTATAT---TTTTC 416
QY 2614 atttgagaattctcaattgtatctagttaaatagcacagtttgaaactgtctgagac 2673
|||||
Db 417 ATTTGAGAATTTCTCAATTGTATCTAGTTAAATAGCACAGTTTGAAACTTGTCTGAGAC 476
QY 2674 tgacttlatcaataatactaaacgcacaagaatcatatccatgtgtatggttagacatt 2733
|||||
Db 477 TGACTTTATCAATTAATCTAAACCGACAAAGATCATATCCATGTGTATGTGTTAGACATT 536
QY 2794 gctgaacagcttcctgacttt 2814
|||||
Db 595 GCTG---CAGTCCGTGACTTT 612

RESULT 15
LOCUS BE617223 749 bp mRNA EST 24-AUG-2000
DEFINITION BE617223 601441906F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846265 5',
mRNA sequence.
ACCESSION BE617223
VERSION BE617223.1 GI:9888161
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 749)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM545 row: k column: 02
High quality sequence stop: 705.

FEATURES

source
1. 749
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846265"
/clone_1lb="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 244 a 166 c 223 g 116 t
ORIGIN

Query Match 16.9%; Score 565.4; DB 110; Length 749;
Best Local Similarity 95.1%; Pred. No. 1.8e-116;
Matches 714; Conservative 0; Mismatches 21; Indels 16; Gaps 12;

QY 677 aagacaagacaggtcgcaaaacagatgtctgccaaa-gcgttgagatcctgcagaa 735
Db 1 AAGACAAGACAGGCTGCGAAACAGATGCTGCTGCCAAAGCGTTGAGGATCCTGCAGAA 60
QY 736 tgagccctgcagagaggtgtgaatgtgaagaatcgaagaagaanaatctcaa 795
Db 61 TGAGCCCTGCGCAGAGAGCTGAGGTGAATGAGAATCCGAAGAAGAAATCTCAA 120
QY 796 taaatctgaataagtcagtgtt-gagattgcaactaaacggaacttgctgtgaatt 854
Db 121 TAAATCTGAATTAAGTCAAGTGTGGAGATTGCACCTTAACGGAACCTTGCTGTGAATT 180
QY 855 tcgaggtgccccgagagagtgccccacacatgaagaacttgtgaccaaggttcgg 914
Db 181 TCGAGGTGGCCCGGAGAGTGGCCACCCACATGAAGAAC-TTGTGACCAAGGTTTCGG 239
QY 915 ttggygaagtttgtgyggaagtgaaggyaaagaagaattcaaaagaaaaatgcg 974
Db 240 TTGGGGAG-TTGTGGGGAGGTGAAGGAAAGCAAGATTTCAAAGAAATGCCG 298
QY 975 ccataagctgttcttgagagctgaagaagtlaccgcccctgctgcagtgaacgagtaa 1034
Db 299 CCATAGCTGTTCTTGAGGAGCTGAAGAAGTTACCG-CCCTGCCTGCAGTTGAACGAGTAA 357
QY 1035 agcctagaatcaaaagaacaaaccatagtcaagccaagccagaatatg 1094
Db 358 AGCCTAGAATCAAAAAGAAACAACCCATAGTCAAGCCACAGACAAGCCCAATATG 417
QY 1095 gccaggygatcaatccgattagccgactgcccagatccagcagycaaaaaagagaag 1154
Db 418 GCCA-GGGATCAATCCGATTACCCGACTGGCCCATCCAGCAGGCAAAAAAGAGAAGG 476
QY 1155 agccagagtacagctcctcaacagagcgaagcctcccgccgcgcagagagttgtatgc 1214
Db 477 AGCCAGAGTACACGCTCCTCACAGAGCGAGG-CTCCCGCGCCGAGGGAG-TTGTGATGC 534
QY 1215 agtgtgaagttgaaacacacatgcagaaggaagcagcaccac-aagaagttgccaag 1273
Db 535 AGGTGAAGTGGGAAACACACTGCAGAAGAACGGGCACCAACAAGAAGGTGGCCAAG 594
QY 1274 cgcaatgcagccgcgagacatgctgagatccttgttt--caaagtcgcgcagcgcgcagc 1331
Db 595 CGCAATGCAGCCGAGAACATGCTGGAGATCCTTGTTTCAAAAGTCCCGCAGGGGACG 654

QY 1332 ccaccaaacccgcgactcaagtcagaggagaagacaccataaagaacccaggggatgaa 1391
Db 655 CAACAAACCGGACTC--AGTCAGAGGAGAGAAGACACACCATAA--GAAACCGGGGATGAA 709
QY 1392 gaaaagtaacctttttgaacctgctctgg 1422
Db 710 GACACGTAACTTTTGGAACTGGGCCGGGG 740

Search completed: April 5, 2001, 00:30:32
Job time: 53128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:31:19 ; Search time 701.87 Seconds
(without alignments)
1791.954 Million cell updates/sec

Title: US-09-316-048-7
Perfect score: 3348
Sequence: 1 actctcgcgcggctgcgqg.....taatactataaaaaaaaaa 3348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /N_Geneseq_36:*/
2: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
3: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
4: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
5: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
6: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
7: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
8: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
9: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
10: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3139.	93.8	3260	21	Z93327	Human stauflen CDNA
2	3076	91.9	3217	21	Z93323	Human stauflen CDNA
3	2987.2	89.2	3190	20	X90786	DNA encoding human
4	2961.6	88.5	3506	21	Z93322	Human stauflen CDNA
5	2926	87.4	3142	21	Z93324	Human stauflen CDNA
6	1255	37.5	2859	21	Z93328	Mouse stauflen CDNA
7	1107.2	33.1	1127	20	X39991	Prostate cancer as
8	798.8	23.9	946	20	X40002	Prostate cancer as
9	755.6	22.6	844	20	X40000	Prostate cancer as
10	753.2	22.5	773	20	X39992	Prostate cancer as
11	551	16.5	825	20	X40001	Prostate cancer as
12	501.4	15.0	794	20	X39993	Prostate cancer as

C	13	433.8	13.0	660	21	Z79981	Human colon cancer
C	14	412.2	12.3	727	21	Z80269	Human colon cancer
C	15	327.8	9.8	708	21	Z80703	Human colon cancer
C	16	286.6	8.6	305	16	T20075	Human gene express
C	17	119.8	3.6	300	20	Z17757	Human gene express
C	18	119.8	3.6	300	20	Z14226	Human gene express
C	19	115.2	3.4	769	20	Z17411	Hereditary haemoch
C	20	88.2	2.6	235033	19	V57926	Hereditary haemoch
C	21	88.2	2.6	237326	19	V57903	Serglycin - proteo
C	22	86.8	2.6	17327	14	Q44278	Human adenosine re
C	23	85.8	2.6	6870	21	A34833	Human adenosine re
C	24	85.8	2.6	8055	21	A34834	Human adenosine re
C	25	85.4	2.6	228	19	V49914	Human Iga nephropa
C	26	85	2.5	13585	17	T11549	Tumour rejection a
C	27	85	2.5	56583	21	A35003	Human adenosine re
C	28	84.8	2.5	1354	19	V35619	Human transcriptio
C	29	84	2.5	84607	20	X90847	Human PACAP genom
C	30	83.8	2.5	9365	21	Z50359	Human SHOX (short
C	31	83.8	2.5	32367	19	V35620	Human secreted pro
C	32	83.4	2.5	2529	21	A16662	Human secreted pro
C	33	83.4	2.5	2688	18	T72165	Human secreted pro
C	34	83.4	2.5	138169	21	A34791	Human adenosine re
C	35	83.4	2.5	141589	21	A35005	Human adenosine re
C	36	83.4	2.5	141589	21	A35030	Human adenosine re
C	37	83.4	2.5	162450	21	Z86967	Retinoblastoma bin
C	38	83.2	2.5	17131	21	Z60888	DNA encoding a hum
C	39	83	2.5	84607	20	X90847	Human PACAP genom
C	40	82.4	2.5	625	19	V35615	SHOX gene exon Vb
C	41	82.2	2.5	300	20	Z12542	Human gene express
C	42	82.2	2.5	1040	20	X79016	Human secreted pro
C	43	82.2	2.5	2170	20	X03027	Human IL-1ra BAC c
C	44	82.2	2.5	3071	17	T11778	Human chromosome-1
C	45	82.2	2.5	11901	20	X02998	Human IL-1ra BAC c

ALIGNMENTS

RESULT 1	
ID Z93327	standard; cDNA; 3260 BP.
XX AC Z93327;	
XX DT 04-JUL-2000	(first entry)
XX DE Human stauflen CDNA.	
XX XX	
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW acquired immune deficiency syndrome; protease; ; human; ss.	
XX XX	
OS Homo sapiens.	
XX XX	
FH Key	Location/Qualifiers
FT CDS	409..1311
FT FT	/*tag= a
FT FT	/product= Stauflen protein
XX XX	
PN CA2238656-A1.	
XX XX	
PD 22-NOV-1999.	
XX XX	
PF 22-MAY-1998;	98CA-2238656.
XX XX	
PR 22-MAY-1998;	98CA-2238656.
XX XX	
PA (UYMO-) UNIV MONTREAL.	
XX XX	
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;	
PI Moulard AJ;	
XX XX	
DR WPI; 2000-246924/22.	

DR P-PSDB; Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1b; 96pp; English.
XX
CC Stauflen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC stauflen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian stauflen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauflen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
SQ Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

Query Match 93.8%; Score 3139; DB 21; Length 3260;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 3259; Conservative 0; Mismatches 0; Indels 90; Gaps 3;

QY 1 acttcctgcgagctgcggcgccctgagcgctctcagcggttgcgcgcggtgcgcgt 60
Db 1 acttcctgcgagctgcggcgccctgagcgctctcagcggttgcgcgcggtgcgcgt 60
QY 61 ctctctcgctcccgctctcctttgaccgcctcccccgcggcccgcgccgcgcctc 120
Db 61 ctctctcgctcccgctctcctttgaccgcctcccccgcggcccgcgccgcgcctc 120
QY 121 ctccagcgccactccgcctcttcctcccttcgtcccttctcctccttttctcct 180
Db 121 ctccagcgccactccgcctcttctcctcctt----- 151
QY 181 ctctcctccctcctcgccgcaacgcgcccagaccgcccggggagcgagtcggagc 240
Db 152 -----agc 154
QY 241 agcagccagcagcagcaggtgagtttgctctgtcgccagcgctgagtgagtggc 300
Db 155 agcagccagcagcagcaggtgagtttgctctgtcgccagcgctgagtgagtggc 214
QY 301 gtgactcgcgtcactgcgaacctccacctcccaaggtcagcgatttcccactcagcctc 360
Db 215 gtgactcgcgtcactgcgaacctccacctcccaaggtcagcgatttcccactcagcctc 274
QY 361 ccgataagctgagattaca-gagtttataaccacttaacctctcagaactgaaacaaga 419
Db 275 ccgataagctgagattacaagaggttataaccacttaacctctcagaactgaaacaaga 334
QY 420 caacatgttctcctggaagccctctttaaataaaagaagcataaaccctactgtagaac 479
Db 335 caacatgttctcctggaagccctctttaaataaaagaagcataaaccctactgtagaac 394
QY 480 taaatgcactgtgcatgaaacttgaaaaaaaccaatgtataagcctgtgacccttact 539
Db 395 taaatgcactgtgcatgaaacttgaaaaaaaccaatgtataagcctgtgacccttact 454
QY 540 ctcgagtgcagtccacctataactacaacatgagagagtgcttatcccccgagtgact 599
Db 455 ctcgagtgcagt---ccataactacaacatgagagagtgcttatcccccgagtgact 511

QY 600 tttaaccattccagttccaccttacttataccaagtggaacttctgtggagcagc 659
Db 512 tttaaccattccagttccaccttacttataccaagtggaacttctgtggagcagc 571
QY 660 aatttaatgcaagaagaagaacaagacagctgcgaacaacagatgctgtgccaaagcgt 719
Db 572 aatttaatgcaagaagaagaacaagacagctgcgaacaacagatgctgtgccaaagcgt 631
QY 720 tgaagatcctgcagaatgagccctgcgcagagagctggaagtgaatgaaagaatccg 779
Db 632 tgaagatcctgcagaatgagccctgcgcagagagctggaagtgaatgaaagaatccg 691
QY 780 aagaagaaaatctcaataaactgaataaagtgtttgagatgcaactaaacyga 839
Db 692 aagaagaaaatctcaataaactgaataaagtgtttgagatgcaactaaacyga 751
QY 840 acttgctgtgaatttcgaaggtgcccgggagagtgcccacccacatgaagaacttg 899
Db 752 acttgctgtgaatttcgaaggtgcccgggagagtgcccacccacatgaagaacttg 811
QY 900 tgaaccaaggttcggttggggagtttgtggggaaagtgaaagggaaaaagcaagaagatt 959
Db 812 tgaaccaaggttcggttggggagtttgtggggaaagtgaaagggaaaaagcaagaagatt 871
QY 960 caaagaaaaatgccgcacatagctgttcttgagagctgaaagaagttaccgcccctgctg 1019
Db 872 caaagaaaaatgccgcacatagctgttcttgagagctgaaagaagttaccgcccctgctg 931
QY 1020 cagttgaaacgagtaagcctagaatcaaaaaagaanaaacccatagtcgaagccacaga 1079
Db 932 cagttgaaacgagtaagcctagaatcaaaaaagaanaaacccatagtcgaagccacaga 991
QY 1080 caagcccaagaatatgcccaggggatccaatccgattagccgactgcccagatccaagcag 1139
Db 992 caagcccaagaatatgcccaggggatccaatccgattagccgactgcccagatccaagcag 1051
QY 1140 caaaaaaggaagaagcagagtaacagctctcaacagagcgagcctccgcgcgcga 1199
Db 1052 caaaaaaggaagaagcagagtaacagctctcaacagagcgagcctccgcgcgcga 1111
QY 1200 gggagttgtgatgcaggtgaaggttggaaccacactgcagaaggaagcgaccacaaca 1259
Db 1112 gggagttgtgatgcaggtgaaggttggaaccacactgcagaaggaagcgaccacaaca 1171
QY 1260 agaaggtggccaagcgcaatgcaagccgagacatgctgagatccttggttcaagtcc 1319
Db 1172 agaaggtggccaagcgcaatgcaagccgagacatgctgagatccttggttcaagtcc 1231
QY 1320 cgcagcggcagcccaaccccgcaactcaagtccagaggaagaagacccataaagaaac 1379
Db 1232 cgcagcggcagcccaaccccgcaactcaagtccagaggaagaagacccataaagaaac 1291
QY 1380 caggggatggaagaagaagtaacctttttgaaacctgctctggygataaagtggacta 1439
Db 1292 caggggatggaagaagaagtaacctttttgaaacctgctctggygataaagtggacta 1351
QY 1440 gtaataaagaagatgagttcaggaatgccttatcctaagtcatcagcagctgcctgtgaa 1499
Db 1352 gtaataaagaagatgagttcaggaatgccttatcctaagtcatcagcagctgcctgtgaa 1411
QY 1500 ttcttcccatgtgcccagaggtgcgcccagcgtgtagaggttagtcaaggacatcacacca 1559
Db 1412 ttcttcccatgtgcccagaggtgcgcccagcgtgtagaggttagtcaaggacatcacacca 1471
QY 1560 aagattttacaggcagctccgaatccttgcgaaggccacggtaacgtgcatgatagcc 1619
Db 1472 aagattttacaggcagctccgaatccttgcgaaggccacggtaacgtgcatgatagcc 1531
QY 1620 gagagttgtgtatgtggggcacctcgcccacagcgagaccatttcaagaataacatct 1679
Db 1532 gagagttgtgtatgtggggcacctcgcccacagcgagaccatttcaagaataacatct 1591
QY 1680 ctcaagccacgtlaccatcctgacctctcaagagaccctctgagcaactgactatcctt 1739

|||||
Db 1592 cttcagccacgtraccccatgtgacctctcagcagacctctgagcaactgactatctt 1651
QY 1740 ccagagtcaggagattccaggttgaatacaaaagacttccccaaaaacaagaacgaat 1799
Db 1652 ccagagtcaggagattccaggttgaatacaaaagacttccccaaaaacaagaacgaat 1711
QY 1800 ttgtatctctatcaatttgtctctctcagccacctctgatacagccatggtatcggcaag 1859
Db 1712 ttgtatctctatcaatttgtctctctcagccacctctgatacagccatggtatcggcaag 1771
QY 1860 atgtggaagtcctgcacatgatatgtgctgcgtgaaacatcttaagtgtgctgtgagtgg 1919
Db 1772 atgtggaagtcctgcacatgatatgtgctgcgtgaaacatcttaagtgtgctgtgagtgg 1831
QY 1920 accaacaagaagtagagatgccaagaacaggaacagcaaatgtctgtgtggaagt 1979
Db 1832 accaacaagaagtagagatgccaagaacaggaacagcaaatgtctgtgtggaagt 1891
QY 1980 gctgaacctttctggccatgaaacattataaaatcccaacatatatactgaaataactg 2039
Db 1892 gctgaacctttctggccatgaaacattataaaatcccaacatatatactgaaataactg 1951
QY 2040 aaactgctttgaaaaatttggaatttctgtatacctccagtgggccgagaagacacggtgg 2099
Db 1952 aaactgctttgaaaaatttggaatttctgtatacctccagtgggccgagaagacacggtgg 2011
QY 2100 aaagatgtggcgacgacgaggaagaaacacaaagaagcggtgtggccgac 2159
Db 2012 aaagatgtggcgacgacgaggaagaaacacaaagaagcggtgtggccgac 2071
QY 2160 tggactgtgtggggttgtgtgtatgagccactcgtgtgacctggtccctacgcaata 2219
Db 2072 tggactgtgtggggttgtgtgtatgagccactcgtgtgacctggtccctacgcaata 2131
QY 2220 gcaagctgctgtggggaagaagggctgcccagccagctggtctcccgggacaccagcag 2279
Db 2132 gcaagctgctgtggggaagaagggctgcccagccagctggtctcccgggacaccagcag 2191
QY 2280 atccacacccctgggcaacctccgtgtttgtgtcttttttccctgtgtgaaagaagaac 2339
Db 2192 atccacacccctgggcaacctccgtgtttgtgtcttttttccctgtgtgaaagaagaac 2251
QY 2340 ggcacgaccccttctcaagctgtgctcaactcagacacatttgggacaaaacctggaagcca 2399
Db 2252 ggcacgaccccttctcaagctgtgctcaactcagacacatttgggacaaaacctggaagcca 2311
QY 2400 tggcagagagagggccttgaacgggccccagagctaaaagcaccagagaaaaatcaaatgct 2459
Db 2312 tggcagagagagggccttgaacgggccccagagctaaaagcaccagagaaaaatcaaatgct 2371
QY 2460 tccctactcagcgtgacccaacttttctagtggtgccaagggccccacacctcctgcaatgac 2519
Db 2372 tccctactcagcgtgacccaacttttctagtggtgccaagggccccacacctcctgcaatgac 2431
QY 2520 ccacaccatcaccaactgtcttctctccaacagtgatctgtattcttagtttcattatt 2579
Db 2432 ccacaccatcaccaactgtcttctctccaacagtgatctgtattcttagtttcattatt 2491
QY 2580 tcttttgattgatagacacataataaaatttcaatttgagaatttctcaattgtatcta 2639
Db 2492 tcttttgattgatagacacataataaaatttcaatttgagaatttctcaattgtatcta 2551
QY 2640 gttaaatagcacagtttggaaaactgtctgaagactgacttatacaataatctaaccgaca 2699
Db 2552 gttaaatagcacagtttggaaaactgtctgaagactgacttatacaataatctaaccgaca 2611
QY 2700 aagatcatatccatgtgtatgtgtgttagacatttattcatatgacaaaccagagacag 2759
Db 2612 aagatcatatccatgtgtatgtgtgttagacatttattcatatgacaaaccagagacag 2671
QY 2760 ttccagtgatgcaaatgtgtgcccctgtgttcagctgaaacagtcctggaacttcaaaa 2819
|||||

Db 2672 ttccagtgatgcaaatgtgtgcccctgtgttcagctgaaacagtcctggaacttcaaaa 2731
QY 2820 accttgaataagtcctccacagtggtataaatgtgacaatttaggaattttaaactttag 2879
Db 2732 accttgaataagtcctccacagtggtataaatgtgacaatttaggaattttaaactttag 2791
QY 2880 atgatacttgtgtccattttatttcaatttttattttttgtttaaagcaaacaggaactaa 2939
Db 2792 atgatacttgtgtccattttatttcaatttttattttttgtttaaagcaaacaggaactaa 2851
QY 2940 atgaactttgatctctgtttttaaagattatttaaaaaacattgtgtatctatataatgac 2999
Db 2852 atgaactttgatctctgtttttaaagattatttaaaaaacattgtgtatctatataatgac 2911
QY 3000 tcttgaggacttagcttccactacactacagagatgatactccatgtagttccatataaac 3059
Db 2912 tcttgaggacttagcttccactacactacagagatgatactccatgtagttccatataaac 2971
QY 3060 ctgcagaagtgatltccagaagtgctcgataactgtttaattacatctccatgaaggctgaaa 3119
Db 2972 ctgcagaagtgatltccagaagtgctcgataactgtttaattacatctccatgaaggctgaaa 3031
QY 3120 agaataacctacggttctgtataacagctgtgtgcttttgatgtgtgttaactgtlacaca 3179
Db 3032 agaataacctacggttctgtataacagctgtgtgcttttgatgtgtgttaactgtlacaca 3091
QY 3180 gaagtggtgacactgagggctcggtgtgtgtcgataggaanaacctggtgaagccctgag 3239
Db 3092 gaagtggtgacactgagggctcggtgtgtgtcgataggaanaacctggtgaagccctgag 3151
QY 3240 ttaagtactgtctccatccattgtttaaagctgtgaatttttctcccatggaatgttaagta 3299
Db 3152 ttaagtactgtctccatccattgtttaaagctgtgaatttttctcccatggaatgttaagta 3211
QY 3300 aaacttaagtggtgtgcatcaataaatggtlaataactaaaaa 3348
Db 3212 aaacttaagtggtgtgcatcaataaatggtlaataactaaaaa 3260

RESULT 2
Z93323
ID Z93323 standard; cDNA; 3217 BP.
XX
AC Z93323;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
OS Homo sapiens.
XX
FH key
FT CDS
FT
FT
FT
CA2238656-A1.
PD 22-NOV-1999.
XX
XX 22-MAY-1998; 98CA-2238656.
PE 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L;
PI Moulard AJ;
XX

Location/Qualifiers
363..1853
/*tag= a
/product= stauflen

DR WPI; 2000-246924/22.
DR P-PSDB; Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96bp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC Z93322.
XX
SQ Sequence 3217.BP; 889 A; 796 C; 742 G; 790 T; 0 other;

Query Match 91.9%; Score 3076; DB 21; Length 3217;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 3217; Conservative 0; Mismatches 0; Indels 131; Gaps 1;

QY 1 acttcctgcccgggctgcyggcgccctgaagcgctctcagcgcttgcyggcggtgcgct 60
DB 1 acttcctgcccgggctgcyggcgccctgaagcgctctcagcgcttgcyggcggtgcgct 60
QY 61 ctctctcgctcccgctctcctttgacgcctcccccccgccgcygcgcgcgcctc 120
DB 61 ctctctcgctcccgctctcctttgacgcctcccccccgccgcygcgcgcgcctc 120
QY 121 ctccacgycgaactccgcctctctcccttccttccttccttccttccttccttcct 180
DB 121 ctccacgycgaactccgcctctctcccttccttccttccttccttccttccttcct 180
QY 181 ctctcctccctcctctgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
DB 181 ctctcctccctcctctgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
QY 241 agcagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
DB 241 agcagc----- 246
QY 301 gtgatctcgctcaactgcaacctcccaactcccaagtcagcgatttcccaactcagcctc 360
DB 247 ----- 246
QY 361 ccgataagctgagattacagagtttaataccaacttaacctcagaactgaaacaaagac 420
DB 247 -----cagagtttaataccaacttaacctcagaactgaaacaaagac 289
QY 421 aacattgttctcgtgaagccctcttttaaaaaaagaagataccctactgtagaact 480
DB 290 aacattgttctcgtgaagccctcttttaaaaaaagaagataccctactgtagaact 349
QY 481 aaatgcactgtgcatgaaacttggaaaaaaaccaatgtataagcctgttgaccctactc 540
DB 350 aaatgcactgtgcatgaaacttggaaaaaaaccaatgtataagcctgttgaccctactc 409
QY 541 tcggaatgcagtcacactataactacaacatgagaggttctatcccccgaggtactt 600
DB 410 tcggaatgcagtcacactataactacaacatgagaggttctatcccccgaggtactt 469

QY 601 ttaccatttcaggttcacacttacttatcaagtgyaactttctgtggagacagca 660
DB 470 ttaccatttcaggttcacacttacttatcaagtgyaactttctgtggagacagca 529
QY 661 attaatgcaaaaggaaagacaaagacagcgtcggaacacagatgctgtcgcaaacggtt 720
DB 530 attaatgcaaaaggaaagacaaagacagcgtcggaacacagatgctgtcgcaaacggtt 589
QY 721 gagatcctgcagaatgagccctgcagagagagcgtggaagtgaatggaagaaatccga 780
DB 590 gagatcctgcagaatgagccctgcagagagagcgtggaagtgaatggaagaaatccga 649
QY 781 agaagaaatcctaataatctgaataagtcagtgattgagattgcacttaacggaa 840
DB 650 agaagaaatcctaataatctgaataagtcagtgattgagattgcacttaacggaa 709
QY 841 ctgcctgtgatttcgaggtgcccgggagagtgcccccacatgaagaacttgt 900
DB 710 ctgcctgtgatttcgaggtgcccgggagagtgcccccacatgaagaacttgt 769
QY 901 gaccaggttcgcggttgaggagtttggggggaagtggaagggaaaaagcaagaatcttc 960
DB 770 gaccaggttcgcggttgaggagtttggggggaagtggaagggaaaaagcaagaatcttc 829
QY 961 aaagaaaatccgcgcataagctgttcttgaggagctgaagaagtaccgcccctgctgc 1020
DB 830 aaagaaaatccgcgcataagctgttcttgaggagctgaagaagtaccgcccctgctgc 889
QY 1021 agtgaacgagtaaaagcctagaatcaaaaagaaaacaaacccatagtcagccacagac 1080
DB 890 agtgaacgagtaaaagcctagaatcaaaaagaaaacaaacccatagtcagccacagac 949
QY 1081 aagccagaatatgcccaggggagatcaatccgattagccgactgcccagatccagacgac 1140
DB 950 aagccagaatatgcccaggggagatcaatccgattagccgactgcccagatccagacgac 1009
QY 1141 aaaaaagagaaggagccagagatcacgctcctcacagagcgaggcctcccgcgcgag 1200
DB 1010 aaaaaagagaaggagccagagatcacgctcctcacagagcgaggcctcccgcgcgag 1069
QY 1201 ggagttgtgatgcaagtggaagcttggaaccacacactgcagaagaaagcggcaccacaa 1260
DB 1070 ggagttgtgatgcaagtggaagcttggaaccacacactgcagaagaaagcggcaccacaa 1129
QY 1261 gaaggtgccaagcgcaatgcagccgagaacatgctgagatccttggttcaaaagctcc 1320
DB 1130 gaaggtgccaagcgcaatgcagccgagaacatgctgagatccttggttcaaaagctcc 1189
QY 1321 gcagcgccagcccaaccaacccgcactcaagtcagagagaagaaacacataaagaacc 1380
DB 1190 gcagcgccagcccaaccaacccgcactcaagtcagagagaagaaacacataaagaacc 1249
QY 1381 agggatggaagaaagtaacctttttgaacctggtctgaggatgaaatgggactag 1440
DB 1250 agggatggaagaaagtaacctttttgaacctggtctgaggatgaaatgggactag 1309
QY 1441 taataagagatgagttcagatgccttatctaagtcatcagcagctgctgtggaat 1500
DB 1310 taataagagatgagttcagatgccttatctaagtcatcagcagctgctgtggaat 1369
QY 1501 tcttcccatgtgccgaggtcgcccaggtgtlagagttagtcgaagacatcacaccaa 1560
DB 1370 tcttcccatgtgccgaggtcgcccaggtgtlagagttagtcgaagacatcacaccaa 1429
QY 1561 agatttaccagggcagctccgaatcctgcgaagcgaagcgaagtaactgcatgatacccg 1620
DB 1430 agatttaccagggcagctccgaatcctgcgaagcgaagcgaagtaactgcatgatacccg 1489
QY 1621 agagttgttgatggggcacctcgcccacagccgagaccaatttaagaataacatctc 1680
DB 1490 agagttgttgatggggcacctcgcccacagccgagaccaatttaagaataacatctc 1549

QY 1681 ttcaaggccaagtaaccatggaacctctcacgagagaccccttgagcaactggaactatctttc 1740
|||||
Db 1550 ttcaaggccaagtaaccatggaacctctcacgagagaccccttgagcaactggaactatctttc 1609
QY 1741 cagaagtcagaaggatcccaaggttgataacaaagactcccccaaaaaacaagaagaaatt 1800
|||||
Db 1610 cagaagtcagaaggatcccaaggttgataacaaagactcccccaaaaaacaagaagaaatt 1669
QY 1801 tgaatctcttaacaaattgctcctctcagccacctctgatacagccatggtatcggaagga 1860
|||||
Db 1670 tgaatctcttaacaaattgctcctctcagccacctctgatacagccatggtatcggaagga 1729
QY 1861 tgtgagagtcctgccaatgatatgctgcgcgtgaacatcttaaaagtgtgtctgagtgga 1920
|||||
Db 1730 tgtgagagtcctgccaatgatatgctgcgcgtgaacatcttaaaagtgtgtctgagtgga 1789
QY 1921 ccaacaagtaacagagatgccaagaacaggaacgaaccaatgtctgtgtggaaggtg 1980
|||||
Db 1790 ccaacaagtaacagagatgccaagaacaggaacgaaccaatgtctgtgtggaaggtg 1849
QY 1981 ctgaacctttcttgccatggaaccattataaaatcccaacataatactgaaatactga 2040
|||||
Db 1850 ctgaacctttcttgccatggaaccattataaaatcccaacataatactgaaatactga 1909
QY 2041 aactgcttgaaaaatttgaaattctgatacctccagtgggccgagaacagtggtgga 2100
|||||
Db 1910 aactgcttgaaaaatttgaaattctgatacctccagtgggccgagaacagtggtgga 1969
QY 2101 aagatgtggcagcagcagcaggaagaacaacagaacaaagaagcgcgctgtgcccgtc 2160
|||||
Db 1970 aagatgtggcagcagcagcaggaagaacaacagaacaaagaagcgcgctgtgcccgtc 2029
QY 2161 ggaactgtgcggggtttgtgtgtagtggccaactcggtgaacctggcggtccctacgcaatag 2220
|||||
Db 2030 ggaactgtgcggggtttgtgtgtagtggccaactcggtgaacctggcggtccctacgcaatag 2089
QY 2221 cagctgccttgggggaagaagggctgcccagccagcgtggttctccgggacaccagcaga 2280
|||||
Db 2090 cagctgccttgggggaagaagggctgcccagccagcgtggttctccgggacaccagcaga 2149
QY 2281 tccacacccctgggcaacctccggtgttggtcttttttccctgtgtgaagaagaacag 2340
|||||
Db 2150 tccacacccctgggcaacctccggtgttggtcttttttccctgtgtgaagaagaacag 2209
QY 2341 gcacgaccccttctcaagctggtcctaactcagacacattgggacaacccctgacagccat 2400
|||||
Db 2210 gcacgaccccttctcaagctggtcctaactcagacacattgggacaacccctgacagccat 2269
QY 2401 gccagagagaaggcctttgacccgccccagagcgtaaaaagcaccagaagaataatgctt 2460
|||||
Db 2270 gccagagagaaggcctttgacccgccccagagcgtaaaaagcaccagaagaataatgctt 2329
QY 2461 cctactcagcgtgacccaactttctctagtggtgccaagcggccccacacactcctgcaatag 2520
|||||
Db 2330 cctactcagcgtgacccaactttctctagtggtgccaagcggccccacacactcctgcaatag 2389
QY 2521 cacaccatcaccaactgcttctctccaacagtgatctgtattcttagtttcattattt 2580
|||||
Db 2390 cacaccatcaccaactgcttctctccaacagtgatctgtattcttagtttcattattt 2449
QY 2581 cttttgattgatatgacactatataataatttcatttgagaatttctcaattgtatatag 2640
|||||
Db 2450 cttttgattgatatgacactatataataatttcatttgagaatttctcaattgtatatag 2509
QY 2641 ttaaatagcaagtttgaaaaactgtctgagagctgactttatcaataatctaacggacaa 2700
|||||
Db 2510 ttaaatagcaagtttgaaaaactgtctgagagctgactttatcaataatctaacggacaa 2569
QY 2701 agatcataatcaatgtgtatgtgtgtagacaattttatctcatgtaactaacccagagcagt 2760
|||||
Db 2570 agatcataatcaatgtgtatgtgtgtagacaattttatctcatgtaactaacccagagcagt 2629
QY 2761 ttcaagtatgcaaaattgtgtgcccctgtgttcaagctgaacagctcctgacttcaaaaa 2820

Db 2630 ttcaagtatgcaaaattgtgtgcccctctggttcaagctgaacagctcctgacttcaaaaa 2689
QY 2821 cctgaataagctcccaacagttgtataaatggaacattttagaattttaactttaga 2880
|||||
Db 2690 cctgaataagctcccaacagttgtataaatggaacattttagaattttaactttaga 2749
QY 2881 tgatcatttggttccatttttatttccatttttatttttggtaaagcaaacaggaacttaa 2940
|||||
Db 2750 tgatcatttggttccatttttatttccatttttatttttggtaaagcaaacaggaacttaa 2809
QY 2941 tgaacttgatcctctgttttaaaagattataaaaaacattgtgtatctatacatatgct 3000
|||||
Db 2810 tgaacttgatcctctgttttaaaagattataaaaaacattgtgtatctatacatatgct 2869
QY 3001 ctgaggaactagccttcaactacactacagagatgatctccatgtagtccatataaacc 3060
|||||
Db 2870 ctgaggaactagccttcaactacactacagagatgatctccatgtagtccatataaacc 2929
QY 3061 tgcagagtgatctccagagtgctcgatactgttaattatcatctccattagggctgaana 3120
|||||
Db 2930 tgcagagtgatctccagagtgctcgatactgttaattatcatctccattagggctgaana 2989
QY 3121 gaatgaacctacgcttctgtataacagctgtgtgttcttgatgtgtgttactgtacacag 3180
|||||
Db 2990 gaatgaacctacgcttctgtataacagctgtgtgttcttgatgtgtgttactgtacacag 3049
QY 3181 aagtggtgcactgaaggctctgcgtgtgtgtccgatatggaanaacctggttagccctgcagt 3240
|||||
Db 3050 aagtggtgcactgaaggctctgcgtgtgtgtccgatatggaanaacctggttagccctgcagt 3109
QY 3241 taagtaactgcttccatcatattgtttacgcgtgaatttctcccatggaatgtgaagtaa 3300
|||||
Db 3110 taagtaactgcttccatcatattgtttacgcgtgaatttctcccatggaatgtgaagtaa 3169
QY 3301 aacttaagtgttgtcatcaataaatggttaataactataaaaaa 3348
|||||
Db 3170 aacttaagtgttgtcatcaataaatggttaataactataaaaaa 3217
RESULT 3
X90786
ID X90786 standard; DNA; 3190 BP.
XX
AC X90786;
DT 13-JAN-2000 (first entry)
XX
DE DNA encoding human staufen protein.
XX
KW Human staufen protein; hStau; protein lysate; lung; kidney; testis;
KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;
KW telomerase assembly; maturation; transport; regulation; interaction;
KW inhibitory agent; antibody; immunoassay; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 345..1835
FT /*tag= a
FT /product= "Human staufen protein"
FT /note= "Binds to telomerase RNA"
FT polyA_signal 3172..3177
FT /*tag= b
XX
PN WO9951255-A1.
XX
PD 14-OCT-1999.
XX
PF 06-APR-1999; 99WO-US07533.
XX
PR 06-APR-1998; 98US-0080783.
XX

PA (UYUO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX Greider CW, Le S;
XX WPI: 1999-620168/53.
DR P-PSDB: Y26335.
XX
PT Human stauften polypeptide useful in methods for identifying telomerase
inhibitors
XX
PS Claim 13; Page 41-42; 50pp; English.
XX
CC The present sequence is an isolated DNA encoding human stauften (hStau)
CC protein. The hStau protein was detected in protein lysates from lung,
CC kidney, testis and ovary, but not in brain and heart. It binds
CC specifically to human telomerase RNA and can complex with the
CC telomerase catalytic subunit. The hStau protein may play a role in
CC telomerase assembly, maturation, transport and regulation. Interaction
CC between hStau and telomerase can be inhibited by administering inhibitory
CC agents to the cell. These telomerase inhibitors can be identified using
CC hStau protein. Antibodies generated against hStau can be used in various
CC immunoassays.
XX
SQ Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;

Query Match 89.2%; Score 2987.2; DB 20; Length 3190;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 3144; Conservative 0; Mismatches 8; Indels 132; Gaps 2;

QY 56 cgcgtctctctcggtccgcgttcctttgacgcgcctccccccggccggcgcgccc 115
DB 38 cgcgtccctctcggtccgcgttcctttgacgcgcctccccccggccggcgcgccc 97
QY 116 gcctctccacggcgaactcgcgccttcctccctccttccttccttccttccttc 175
DB 98 gcctctccacggcgaactcgcgccttcctccctccttccttccttccttccttc 157
QY 176 tcctctctctccctcctcctcgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgc 235
DB 158 tcctctctctccctcctcctcgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgc 217
QY 236 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 295
DB 218 ggaagcagcagc----- 228
QY 296 gtgcgctgctctcggtcgaactccacctccacctccacctccacctccacctcc 355
DB 229 ----- 228
QY 356 gcctcccgataagctgagattacagagttatttaaccacttaacctctcagaactga 415
DB 229 -----cagaagttatttaaccacttaacctctcagaactga 266
QY 416 aagacaacatgttccctggaacgccccttlttaaaaaaagaagacataaccctactg 475
DB 267 aagacaacatgttccctggaacgccccttlttaaaaaaagaagacataaccctactg 326
QY 476 gaactaaatgcaactgtgcatgaaacttgaaaaaaacaaatgtataagcctgttgacc 535
DB 327 gaactaaatgcaactgtgcatgaaacttgaaaaaaacaaatgtataagcctgttgacc 386
QY 536 tactctcgatgcagtcacactataactacaacatgagagaggtgtctatcccccgagg 595
DB 387 tactctcgatgcagtcacactataactacaacatgagagaggtgtctatcccccgagg 446
QY 596 tacttttaaccatttcaggtccaccttacttatacaagtgaacttctctgtggagga 655
DB 447 tacttttaaccatttcaggtccaccttacttatacaagtgaacttctctgtggagga 506
QY 656 cagcaatttaatgcaagaagaagaagacagagcgtcgaaaaacagatgtgtgcca 715
DB 507 cagcaatttaatgcaagaagaagaagacagagcgtcgaaaaacagatgtgtgcca 566

QY 716 gcgttgagatccttcgaagaatgagccccctgcagagagcgtggaagtgaatggaagaa 775
DB 567 gcgttgagatccttcgaagaatgagccccctgcagagagcgtggaagtgaatggaagaa 626
QY 776 tccgaagaagaataatcacaataatcgaataaagtcaagtgttgagatgcacttaa 835
DB 627 tccgaagaagaataatcacaataatcgaataaagtcaagtgttgagatgcacttaa 686
QY 836 cgaactgtccttgaaatttcgaagtgcccgaggagagtgccccacacatgaagaac 895
DB 687 cgaactgtccttgaaatttcgaagtgcccgaggagagtgccccacacatgaagaac 746
QY 896 ttgtgaccaaagtttcggttgaggagttgttgggggaagtgaagggaagaagaag 955
DB 747 ttgtgaccaaagtttcggttgaggagttgttgggggaagtgaagggaagaagaag 806
QY 956 attcaagaataaattgcgcacatagctgttcttgaggagctgaagaagttaaccgccc 1015
DB 807 attcaagaataaattgcgcacatagctgttcttgaggagctgaagaagttaaccgccc 866
QY 1016 cctgcagttgaaagagtaaaagcctagaatacaaaaagaataaaacccatagtaagcca 1075
DB 867 cctgcagttgaaagagtaaaagcctagaatacaaaaagaataaaacccatagtaagcca 926
QY 1076 cagacaagcccaagaataatgcccaggagatcaatccgattagccgactgcccagatccag 1135
DB 927 cagacaagcccaagaataatgcccaggagatcaatccgattagccgactgcccagatccag 986
QY 1136 cagcgaataaagagaagagagcagagtaacagctcctcaagagcagagcctccgcgc 1195
DB 987 cagcgaataaagagaagagagcagagtaacagctcctcaagagcagagcctccgcgc 1046
QY 1196 cgcagggagtttgtatgcaggtgaaggttgaaacacacatgcagaagaagcgggcacc 1255
DB 1047 cgcagggagtttgtatgcaggtgaaggttgaaacacacatgcagaagaagcgggcacc 1106
QY 1256 aacaagaagtggccaagcgaatgcagccgagaacatgtctggagatccttggttc 1315
DB 1107 aacaagaagtggccaagcgaatgcagccgagaacatgtctggagatccttggttc 1166
QY 1316 gtcccgagcggcagcccaacacccgcaactcaagtcagaaggagaacacccataaag 1375
DB 1167 gtcccgagcggcagcccaacacccgcaactcaagtcagaaggagaacacccataaag 1226
QY 1376 aaaccagggatggaagaaagtaaccttlttgaacctgctctggygatgaataatgg 1435
DB 1227 aaaccagggatggaagaaagtaaccttlttgaacctgctctggygatgaataatgg 1286
QY 1436 actagtaataaagagagtagtcaagtagccttatctaagtcatacagcagctgcgtct 1495
DB 1287 actagtaataaagagagtagtcaagtagccttatctaagtcatacagcagctgcgtct 1346
QY 1496 ggaattcttcccatgtgtgcccgaggtcgcccgagctgttaggagttagtaagagatacac 1555
DB 1347 ggaattcttcccatgtgtgcccgaggtcgcccgagctgttaggagttagtaagagatacac 1406
QY 1556 accaaagattttacagggcagctccgaatcctctgccaagggcaacgtaactgcatgata 1615
DB 1407 accaaagattttacagggcagctccgaatcctctgccaagggcaacgtaactgcatgata 1466
QY 1616 gcccgagagttgtatggyggcacctcgccacagccgagacattttaagaataaac 1675
DB 1467 gcccgagagttgtatggyggcacctcgccacagccgagacattttaagaataaac 1526
QY 1676 atctctcagggccacgtaccccatgacctctcagagagccctctgagcaactgactat 1735
DB 1527 atctctcagggccacgtaccccatgacctctcagagagccctctgagcaactgactat 1586
QY 1736 ctttcagagtcacaggaattccaggttgatacaaaagacttccccaaaaacaacaagaac 1795
DB 1587 ctttcagagtcacaggaattccaggttgatacaaaagacttccccaaaaacaacaagaac 1646

QY 1796 gaattgtatctcttatcaattgtctcctcagccacctgtatcagccatgtgtatcggc 1855
|||||
Db 1647 gaattgtatctcttatcaattgtctcctcagccacctgtatcagccatgtgtatcggc 1706
QY 1856 aagatgtgtgagtcctgtccatgatatagtcgctgcgtgaaacatcttaagtgtgtctgag 1915
|||||
Db 1707 aagatgtgtgagtcctgtccatgatatagtcgctgcgtgaaacatcttaagtgtgtctgag 1766
QY 1916 ttggaccaacaagaatgacagaatgccaagaacggaacaaatgtctgtgtggg 1975
|||||
Db 1767 ttggaccaacaagaatgacagaatgccaagaacggaacaaatgtctgtgtggg 1826
QY 1976 aggtgtgaaacctttctgtgccaatgaaacatatataactgtaaat 2035
|||||
Db 1827 aggtgtgaaacctttctgtgccaatgaaacatatataactgtaaat 1886
QY 2036 actgaaactgtcttgaaaatttgaaatttctgatacctccagctggcgagagacacggt 2095
|||||
Db 1887 actgaaactgtcttgaaaatttgaaatttctgatacctccagctggcgagagacacggt 1946
QY 2096 gggtaaaagatgttggcagcagcagggaagacaaacaaaggaagcggtgtggc 2155
|||||
Db 1947 gggtaaaagatgttggcagcagcagggaagacaaacaaaggaagcggtgtggc 2006
QY 2156 c-ggctgactgtgtggggttgtgtgtatggtgccaactcgtgacctggcggtccctacg 2214
|
Db 2007 cggtgtgactgtgtggggttgtgtgtatggtgccaactcgtgacctggcggtccctacg 2066
QY 2215 caatagcagctgctgtggggaagaagggctggccacgacgctgttctccgggacacc 2274
|||||
Db 2067 caatagcagctgctgtggggaagaagggctggccacgacgctgttctccgggacacc 2126
QY 2275 agcagatccacacccctgggcacctccggtgttggctttttccctgtgtgaaagaa 2334
|||||
Db 2127 agcagatccacacccctgggcacctccggtgttggctttttccctgtgtgaaagaa 2186
QY 2335 gaaacggcagcagcccttctcgaagctggtcactcagacacattgggacaaacccctggac 2394
|||||
Db 2187 gaaacggcagcagcccttctcgaagctggtcactcagacacattgggacaaacccctggac 2246
QY 2395 agccatgccagagagagcctttgacggcccccagactaaaagcacccagagaaaatcaa 2454
|||||
Db 2247 agccatgccagagagagcctttgacggcccccagactaaaagcacccagagaaaatcaa 2306
QY 2455 atgcttctactcagcgtgacccaactttctagtgtgccaagcggcccccacctctctgc 2514
|||||
Db 2307 atgcttctactcagcgtgacccaactttctagtgtgccaagcggcccccacctctctgc 2366
QY 2515 agtaaccacacacatcaccaactgttctcttccaacagtgatctgtattcttagttcat 2574
|||||
Db 2367 agtaaccacacacatcaccaactgttctcttccaacagtgatctgtattcttagttcat 2426
QY 2575 tatttcttttgattgatatagacactataataaaatttcatttgagaatttctcaattgt 2634
|||||
Db 2427 tatttcttttgattgatatagacactataataaaatttcatttgagaatttctcaattgt 2486
QY 2635 atctaagttaaatagcacagtttgaaaactgtctgagactgacttlatcaataatctaac 2694
|||||
Db 2487 atctaagttaaatagcacagtttgaaaactgtctgagactgacttlatcaataatctaac 2546
QY 2695 cgacaaagatcatatccatgtgtatgtgttagacatttattcatgtgactaacccag 2754
|||||
Db 2547 cgacaaagatcatatccatgtgtatgtgttagacatttattcatgtgactaacccag 2606
QY 2755 gacagtttcagtgatgcaaatgtgtgcccctgtgttcagctgaaacagtcctgacttt 2814
|||||
Db 2607 gacagtttcagtgatgcaaatgtgtgcccctgtgttcagctgaaacagtcctgacttt 2666
QY 2815 caaaaacctgataaagtctccacacagttgtataaaattggacaatttagaattttaaac 2874
|||||
Db 2667 caaaaacctgataaagtctccacacagttgtataaaattggacaatttagaattttaaac 2726
QY 2875 tttagatgatcatattgtgtccatttttatttcatattttttgtttaaagcaaacagga 2934

|||||
Db 2727 tttagatgatcatttggttccatttatttcatatttatttltgttaatgcaaacagga 2786
QY 2935 cttaaatgaactttgatctctgttttaagaattatataaaaaacattgtgtatctatacat 2994
|||||
Db 2787 cttaaatgaactttgatctctgttttaagaattatataaaaaacattgtgtatctatacat 2846
QY 2995 atgctcttgaggacttagctttcactacactacagagatatgatctccatgtgtccata 3054
|||||
Db 2847 atgctcttgaggacttagctttcactacactacactacagagatatgatctccatgtgtccata 2906
QY 3055 taaactgcagagtgatttccagagtgctcgatactgtttaattacatctccatagggc 3114
|||||
Db 2907 taaactgcagagtgatttccagagtgctcgatactgtttaattacatctccatagggc 2966
QY 3115 tgaagaatgacctacggttctgtatatacagctgtgtgtcttltgatgtgtgttactgt 3174
|||||
Db 2967 tgaagaatgacctacggttctgtatatacagctgtgtgtcttltgatgtgtgttactgt 3026
QY 3175 acacagaagtgtgtcactgaggtctgtgtgtgtccgtatggaaacctgttagccct 3234
|||||
Db 3027 acacagaagtgtgtcactgaggtctgtgtgtgtccgtatggaaacctgttagccct 3086
QY 3235 gcgagttaagtactgtctccatcattgtttaaagctggaatttctcccatggaatgt 3294
|||||
Db 3087 gcgagttaagtactgtctccatcattgtttaaagctggaatttctcccatggaatgt 3146
QY 3295 aagtaaaacttaagtggttgcatacaataaatgtaataactaaa 3338
|||||
Db 3147 aagtaaaacttaagtggttgcatacaataaatgtaataactaaa 3190

RESULT 4
Z93322
ID Z93322 standard; cDNA; 3506 BP.
XX
AC Z93322;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH
FH Key Location/Qualifiers
FT CDS 409..2142
FT /*tag= a
FT /product= Stauflen
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83023.
XX
XX Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PR for treating retroviral infections especially human immunodeficiency
PR virus infections (HIV)
XX

PS Claim 4; Figure 1a; 96bp; English.
XX Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

Query Match 88.5%; Score 2961.6; DB 21; Length 3506;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 3224; Conservative 0; Mismatches 124; Indels 158; Gaps 3;

QY 1 acttcctgcccggctgcggcgcctgagcgtctctcagcgtttcgcgcgcgctgcgcgt 60
DB 1 acttcctgcccggctgcggcgcctgagcgtctctcagcgtttcgcgcgcgctgcgcgt 60
QY 61 ctctctcgctcccgctctcctttgacgcgtcccccccgcgccgcgcgcgccgctc 120
DB 61 ctctctcgctcccgctctcctttgacgcgtcccccccgcgccgcgcgcgccgctc 120
QY 121 ctccacggccactccgctctctccctcccttcgtcccttctcctccttttctcct 180
DB 121 ctccacggccactccgctctctccctcccttcgtcccttctcctccttttctcct 180
QY 181 ctctcctccctctctcgccgcacccgcgcagccgcccggcggaagcagtcgcgagc 240
DB 181 ctctcctccctctctcgccgcacccgcgcagccgcccggcggaagcagtcgcgagc 240
QY 241 agcagccag----- 249
DB 241 agcagccaggtttaaccacttaacctctcagaactgaaacaaacacattgttcc 300
QY 250 -----cagcagccaggtgaggtttgtcct----- 275
DB 301 tgyaacgcctctttttaaaaaagtagaactttagacttcatagcactgaaatlaacctg 360
QY 276 -----gtcgccagcgctgagtg 293
DB 361 cactgaaagctgtttacctgcatltgttcaactttgttgaagtgaccatgtctcaagt 420
QY 294 cagtgcgtgatctcgcgtcactgcaacctccacctccacggtcagcgatttcccaact 353
DB 421 caagtgcaagttcagaaccacatctgctgtctctcagggagccaaataactgaaacgaac 480
QY 354 cagcctcccgataagctgagattacagagttatlaaccact----- 395
DB 481 cagtcctctctcacaagcctltgatgatatccttctactactagctctcgtccctct 540
QY 396 -----taacctctcagaactgacaacaaagacacacattgtcctggaagcgct 442
DB 541 gaaaatgcagtgtagcccatccaacactcgtcttaccctctgcacatctatatccacc 600
QY 443 cttttaaaaaaagcataaacccctactgtagaactaatatgcactgtgcatgaaact 502
DB 601 agtgacgtgcagaaagcataaacccctactgtagaactaatatgcactgtgcatgaaact 660
QY 503 ggaataaaacaaatgtataagcctgttgacccttactctcgatgcagtcacactataac 562
DB 661 ggaataaaacaaatgtataagcctgttgacccttactctcgatgcagtcacactataac 720

QY 563 taacaatgagagaggtgtcttatcccccgaggtactttaccactttccagttccacct 622
DB 721 taacaatgagagaggtgtcttatcccccgaggtactttaccactttccagttccacct 780
QY 623 ttacttatacaagtgaacttctgttgaggagacagcaatttaatgccaagaaagaca 682
DB 781 ttacttatacaagtgaacttctgttgaggagacagcaatttaatgccaagaaagaca 840
QY 683 agacagctgcgaaacaacgatgtctgtgcgaaagcgttgagatcctgcagaatgagccc 742
DB 841 agacagctgcgaaacaacgatgtctgtgcgaaagcgttgagatcctgcagaatgagccc 900
QY 743 ctgcagagagcgtggaagtgtgaatggaagagaatccgaagaagaataatcacaataatct 802
DB 901 ctgcagagagcgtggaagtgtgaatggaagagaatccgaagaagaataatcacaataatct 960
QY 803 gaataaagtcaagtgtttgagatgtcacttaaacggaacttgctgtgtaatttcgagtg 862
DB 961 gaataaagtcaagtgtttgagatgtcacttaaacggaacttgctgtgtaatttcgagtg 1020
QY 863 gcccgagaggtgcccacccacacatgaagaactltgtgaccaaaggttcgttgaggag 922
DB 1021 gcccgagaggtgcccacccacacatgaagaactltgtgaccaaaggttcgttgaggag 1080
QY 923 ttgtgagggaagtgtgaagggaagaagattccaagaanaaatgcccacatagct 982
DB 1081 ttgtgagggaagtgtgaagggaagaagattccaagaanaaatgcccacatagct 1140
QY 983 gttcttgagagctgtgaagaagttaccgcccctgctgcagttgacgagtaagcctaga 1042
DB 1141 gttcttgagagctgtgaagaagttaccgcccctgctgcagttgacgagtaagcctaga 1200
QY 1043 atcaaaaagaanaaacaacccatagtcagaagccacagacaagcccagaatatgtgccaggg 1102
DB 1201 atcaaaaagaanaaacaacccatagtcagaagccacagacaagcccagaatatgtgccaggg 1260
QY 1103 atcaatccgattagccgacttggcccagatccagcagcgcaaaaaagagaagcagag 1162
DB 1261 atcaatccgattagccgacttggcccagatccagcagcgcaaaaaagagaagcagag 1320
QY 1163 taacagctctcacaagagcgaagcctcccgcgccgagggagttgtgacgagtgaaag 1222
DB 1321 taacagctctcacaagagcgaagcctcccgcgccgagggagttgtgacgagtgaaag 1380
QY 1223 gtggaaccacactgcagaaggaagcgggcaaccaagaaggttgccaagcgcaatgca 1282
DB 1381 gtggaaccacactgcagaaggaagcgggcaaccaagaaggttgccaagcgcaatgca 1440
QY 1283 gccgagaacatgtctgagatccttgtttcaaaagtcocgcagcgcgccaccaaaccc 1342
DB 1441 gccgagaacatgtctgagatccttgtttcaaaagtcocgcagcgcgccaccaaaccc 1500
QY 1343 gcaactcaagtcagaaggagaagacacccataaagaanaaccagggatggaagaaagtaacc 1402
DB 1501 gcaactcaagtcagaaggagaagacacccataaagaanaaccagggatggaagaaagtaacc 1560
QY 1403 tttttgaacctggtctgaggatggaatggaactagtaataaagagatgagttcagg 1462
DB 1561 tttttgaacctggtctgaggatggaatggaactagtaataaagagatgagttcagg 1620
QY 1463 atgccttatctaagtcatacagcagctgctgtggaattcttcccatggtgccgaggtc 1522
DB 1621 atgccttatctaagtcatacagcagctgctgtggaattcttcccatggtgccgaggtc 1680
QY 1523 gcccaagctgtagaggttagtcaagagacatcacaccaagaatttaaccaggcgagctccg 1582
DB 1681 gcccaagctgtagaggttagtcaagagacatcacaccaagaatttaaccaggcgagctccg 1740
QY 1583 aatcctgccaagcgcaacgtaactgccaatgataagccgagaggtgtgtataggggcacc 1642
DB 1741 aatcctgccaagcgcaacgtaactgccaatgataagccgagaggtgtgtataggggcacc 1800
QY 1643 tcgccacacagccgagaccattttaagaataacatctcttcagggccaagtcaccccatgga 1702

Db 1801 |tcgccacagccgagaccattttaagaataacatctctcagggccacgtaecccatgga 1860
QY 1703 |ccttcacgagagaccctctgagcaactgagactatcttccagagtgccaggatccaggtt 1762
Db 1861 |ccttcacgagagaccctctgagcaactgagactatcttccagagtgccaggatccaggtt 1920
QY 1763 |gaatacaaaagacttccccaaaaacaacaagaacgaatttgatatctctatcaattgtcc 1822
Db 1921 |gaatacaaaagacttccccaaaaacaacaagaacgaatttgatatctctatcaattgtcc 1980
QY 1823 |tctcaagccacctctgatacagccaatggtatcggcaaggaatgtagagtcctgccaatgatg 1882
Db 1981 |tctcaagccacctctgatacagccaatggtatcggcaaggaatgtagagtcctgccaatgatg 2040
QY 1883 |gctgcgctgaacatcttaaaattgtctgagttaggaccaacaagaatgacagatgcca 1942
Db 2041 |gctgcgctgaacatcttaaaattgtctgagttaggaccaacaagaatgacagatgcca 2100
QY 1943 |agacagggaaaacgagaccaatgtctgtgtgtggaggtgctgaaccttttctggccatgaa 2002
Db 2101 |agacagggaaaacgagaccaatgtctgtgtgtggaggtgctgaaccttttctggccatgaa 2160
QY 2003 |ccattataaaatcccaacatatatactgaaaaatactgaaaactgctttgaaaaatttggaa 2062
Db 2161 |ccattataaaatcccaacatatatactgaaaaatactgaaaactgctttgaaaaatttggaa 2220
QY 2063 |ttctgatacctccagtgtagccgagagacaacggtggtaaaagatgtggcagcagcaggg 2122
Db 2221 |ttctgatacctccagtgtagccgagagacaacggtggtaaaagatgtggcagcagcaggg 2280
QY 2123 |aagacaacagaaaacacacaagagcgcggtgtggccggcctggaactgtgctggggttgtgt 2182
Db 2281 |aagacaacagaaaacacacaagagcgcggtgtggccggcctggaactgtgctggggttgtgt 2340
QY 2183 |gatggccaactcggtgacgtggcggtccctaagcaatagcagctgcctgtggggagaagaag 2242
Db 2341 |gatggccaactcggtgacgtggcggtccctaagcaatagcagctgcctgtggggagaagaag 2400
QY 2243 |gctgccacgccagctgtgttctcccgggacaccagcagatccacacccctggcactccgt 2302
Db 2401 |gctgccacgccagctgtgttctcccgggacaccagcagatccacacccctggcactccgt 2460
QY 2303 |gttcggctcttttttccctgtgtgaaagaagaacgycacgaccccttctcaagctgg 2362
Db 2461 |gttcggctcttttttccctgtgtgaaagaagaacgycacgaccccttctcaagctgg 2520
QY 2363 |ctcactcagacacattgggacaaaacctgagacagccatgccaagagagagcctttgaccg 2422
Db 2521 |ctcactcagacacattgggacaaaacctgagacagccatgccaagagagagcctttgaccg 2580
QY 2423 |gccccagagctlaaaagcacaccagagaaaatcaaatgtctcctactcagcgtgacccaact 2482
Db 2581 |gccccagagctlaaaagcacaccagagaaaatcaaatgtctcctactcagcgtgacccaact 2640
QY 2483 |ttctaagtgcacagcgccccaccactcctgcagtaaccaacaccatcacccactgttct 2542
Db 2641 |ttctaagtgcacagcgccccaccactcctgcagtaaccaacaccatcacccactgttct 2700
QY 2543 |cttccaacagtgatctgtatatctcttagtttcatatttcttcttgattgatatgacatat 2602
Db 2701 |cttccaacagtgatctgtatatctcttagtttcatatttcttcttggattgatatgacatat 2760
QY 2603 |ataaaatttcatatttgagaatttctcaattgtatctagtttaaatagcacagtttggaaac 2662
Db 2761 |ataaaatttcatatttgagaatttctcaattgtatctagtttaaatagcacagtttggaaac 2820
QY 2663 |ttgtctgagacactgacttatcaataatctaacgcgacaagaatcatatccatgtgtatgtg 2722
Db 2821 |ttgtctgagacactgacttatcaataatctaacgcgacaagaatcatatccatgtgtatgtg 2880
QY 2723 |gttagacattttatatttcatgtgactaacccagagacagtttcagtgatgcaaatgtgtgc 2782

Db 2881 |gtagacatttttatttcatgtgactaacccagagacagtttcagtgatgcaaatgtgtgc 2940
QY 2783 |cctctgttcagctgaaacagctcctggaacttccaacaaaccttgaataagtcctccacagt 2842
Db 2941 |cctctgttcagctgaaacagctcctggaacttccaacaaaccttgaataagtcctccacagt 3000
QY 2843 |tgtataaattggacaatttaggaattttaaactttagatgatcatlttggttccatttta 2902
Db 3001 |tgtataaattggacaatttaggaattttaaactttagatgatcatlttggttccatttta 3060
QY 2903 |ttcatattttatttttgtttaatgcaaacagagacttaaatgaactttgatctctgttttaa 2962
Db 3061 |ttcatattttatttttgtttaatgcaaacagagacttaaatgaactttgatctctgttttaa 3120
QY 2963 |agattataaaaaacatgtgtatctatacatatggtcctttgaggaacttagcttcaacta 3022
Db 3121 |agattataaaaaacatgtgtatctatacatatggtcctttgaggaacttagcttcaacta 3180
QY 3023 |cactacagatalgatctccatgtagtccatatataaacctgcagagtgatlttccagagtg 3082
Db 3181 |cactacagatalgatctccatgtagtccatatataaacctgcagagtgatlttccagagtg 3240
QY 3083 |ctcgatactgttaattacatctccattaggcgctgaaaagaatgacctacgttctgtata 3142
Db 3241 |ctcgatactgttaattacatctccattaggcgctgaaaagaatgacctacgttctgtata 3300
QY 3143 |cagctgtgttgccttttgatgtgtgttactgtacacagaagtgtgtgcactgaggtctgt 3202
Db 3301 |cagctgtgttgccttttgatgtgtgttactgtacacagaagtgtgtgcactgaggtctgt 3360
QY 3203 |cgtgtgtccgtatgaaaaacctgtgtagccctgcgagtttaagtactgcttccatcatgtg 3262
Db 3361 |cgtgtgtccgtatgaaaaacctgtgtagccctgcgagtttaagtactgcttccatcatgtg 3420
QY 3263 |ttaagctggaaatttctcccatggaatgtgaagtaaaacttaagtgttgatcaaat 3322
Db 3421 |ttaagctggaaatttctcccatggaatgtgaagtaaaacttaagtgttgatcaaat 3480
QY 3323 |aaatgtaataactlaaaaaaaaaa 3348
Db 3481 |aaatgtaataactlaaaaaaaaaa 3506

RESULT 5
Z93324
ID Z93324 standard; cDNA; 3142 BP.
XX
AC Z93324;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human staufen cDNA.
XX
XX Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT CDS 288..1193
FT /*tag= a
FT /product= Staufen protein
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.

QY 1621 agagttgtgtatgagggcacctcgccacagccgagaccattttaaagaataacatctc 1680
|||||
Db 1415 agaagtcgtgtatgagggcacctcgccacagccgagaccattttaaagaataacatctc 1474
QY 1681 ttcagggcacgtaccccatggaacctctcacgagaccctctgagcaactgactatcttc 1740
|||||
Db 1475 ttcagggcacgtaccccatggaacctctcacgagaccctctgagcaactgactatcttc 1534
QY 1741 cagagtcacaggattccaggttgaatacaaaagcttccccaaaaacaacaagaagcaatt 1800
|||||
Db 1535 cagagtcacaggattccaggttgaatacaaaagcttccccaaaaacaacaagaagcaatt 1594
QY 1801 tgtatctcttatcaattgctcctctcacgccaacctctgatacagccatggtatcggcaagga 1860
|||||
Db 1595 tgtatctcttatcaattgctcctctcacgccaacctctgatacagccatggtatcggcaagga 1654
QY 1861 tgtggaagtcctcgccatgatatgctgcgctggaacatcttaagctgctctgagttgga 1920
|||||
Db 1655 tgtggaagtcctcgccatgatatgctgcgctggaacatcttaagctgctctgagttgga 1714
QY 1921 ccaacaaagtlacagagatgccaagaacagaagaacgaaccaatgtctgtgtgaggagt 1980
|||||
Db 1715 ccaacaaagtlacagagatgccaagaacagaagaacgaaccaatgtctgtgtgaggagt 1774
QY 1981 ctgaaacctttctggcacatgaaccatlataaaatcccaacatatatactgaaaatactga 2040
|||||
Db 1775 ctgaaacctttctggcacatgaaccatlataaaatcccaacatatatactgaaaatactga 1834
QY 2041 aactgctttgaaaatttggaaatttctgtataacctccagtgggccgagagacacggtggta 2100
|||||
Db 1835 aactgctttgaaaatttggaaatttctgtataacctccagtgggccgagagacacggtggta 1894
QY 2101 aaggatgttggcagcagcagggaaggaacacaacagaagaagcgctgttggccggt 2160
|||||
Db 1895 aaggatgttggcagcagcagggaaggaacacaacagaagaagcgctgttggccggt 1954
QY 2161 ggaactgttctggggttctgtgtatgagccaactcggtgaacctgtgcccctacgcgaatag 2220
|||||
Db 1955 ggaactgttctggggttctgtgtatgagccaactcggtgaacctgtgcccctacgcgaatag 2014
QY 2221 cagctgcctgtggggaagaaggcctgcccagccaagctgttctcccgggacaccagcaga 2280
|||||
Db 2015 cagctgcctgtggggaagaaggcctgcccagccaagctgttctcccgggacaccagcaga 2074
QY 2281 tccacaccctgggcacctcggtgtgtctttttcccccgtgtgtgaagaagaagaacg 2340
|||||
Db 2075 tccacaccctgggcacctcggtgtgtctttttcccccgtgtgtgaagaagaagaacg 2134
QY 2341 gcaagcacccttctcaagctggctcaactcagacacacattgggacaaaccttgacagccat 2400
|||||
Db 2135 gcaagcacccttctcaagctggctcaactcagacacacattgggacaaaccttgacagccat 2194
QY 2401 gccagagagaggcctttagaccggccccagagctaaaaagcaccagagaagaataatgctt 2460
|||||
Db 2195 gccagagagaggcctttagaccggccccagagctaaaaagcaccagagaagaataatgctt 2254
QY 2461 cctactcagcgtgacccaacttcttagtgtgccaagggccccaccacctctgcagtlacc 2520
|||||
Db 2255 cctactcagcgtgacccaacttcttagtgtgccaagggccccaccacctctgcagtlacc 2314
QY 2521 cacaccatcaccaactgtcttctctccaacagtgatctgtatcttagtttcattattt 2580
|||||
Db 2315 cacaccatcaccaactgtcttctctccaacagtgatctgtatcttagtttcattattt 2374
QY 2581 ctttgattgatatgacactataataaatttccatttgagaatttctcaattgtatctag 2640
|||||
Db 2375 ctttgattgatatgacactataataaatttccatttgagaatttctcaattgtatctag 2434
QY 2641 ttaaatagcacagtttggaactgtctgagactgactttatcaataatcaaccgacaa 2700
|||||
Db 2435 ttaaatagcacagtttggaactgtctgagactgactttatcaataatcaaccgacaa 2494
QY 2701 agatcataatcatgtgtatgtggttagacattttlatctcattgactaaccagagcagt 2760

Db 2495 agatcataatccatgtgtatgtggttagacatttttatttcatttgactaaccagagcagt 2554
QY 2761 ttcagtgtatgcaaatgtgtgcccctctggttccagctgaaacagctcctgactttcaaaa 2820
|||||
Db 2555 ttcagtgtatgcaaatgtgtgcccctctggttccagctgaaacagctcctgactttcaaaa 2614
QY 2821 ccttgaataagctctcccacagttgtataaatgtgacaatttaggaattttaaaccttaga 2880
|||||
Db 2615 ccttgaataagctctcccacagttgtataaatgtgacaatttaggaattttaaaccttaga 2674
QY 2881 tgatcaattggtttccatttttatttcaatttttatttttgglttaatgcaaacaggaacttaa 2940
|||||
Db 2675 tgatcaattggtttccatttttatttcaatttttatttttgglttaatgcaaacaggaacttaa 2734
QY 2941 tgaacttgaatcctgtttttaaagattataaaaaacatgtgtatctatataatggt 3000
|||||
Db 2735 tgaacttgaatcctgtttttaaagattataaaaaacatgtgtatctatataatggt 2794
QY 3001 cttagagacttagctttccactacactacagagatatgatctccatgtagttccatataaac 3060
|||||
Db 2795 cttagagacttagctttccactacactacagagatatgatctccatgtagttccatataaac 2854
QY 3061 tgcagaagtatttccagagtgctcgatactgtttaattacatctccattagggctgaaaa 3120
|||||
Db 2855 tgcagaagtatttccagagtgctcgatactgtttaattacatctccattagggctgaaaa 2914
QY 3121 gaatgaacctacglttctgtatacagctgtgtgctttagatgtgtgtactgtacacag 3180
|||||
Db 2915 gaatgaacctacglttctgtatacagctgtgtgctttagatgtgtgtactgtacacag 2974
QY 3181 aagtggtgacactgaggtctgcgtgtgtgtccgtatggaaaacctgttagccctgcgagt 3240
|||||
Db 2975 aagtggtgacactgaggtctgcgtgtgtgtccgtatggaaaacctgttagccctgcgagt 3034
QY 3241 taagtactgcttccatttcattgttttaacgctggaatttttcccccattggaatgaa 3300
|||||
Db 3035 taagtactgcttccatttcattgttttaacgctggaatttttcccccattggaatgaa 3094
QY 3301 aacttaagtgtttgtcatcaataaatgtataactataaaaaaa 3348
|||||
Db 3095 aacttaagtgtttgtcatcaataaatgtataactataaaaaaa 3142

RESULT 6
Z93328
ID Z93328 standard; cDNA; 2859 BP.
XX
AC Z93328;
XX
DT 04-JUL-2000 (first entry)
XX
DE Mouse stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; mouse; ss.
OS Mus musculus.
XX
FH Key location/Qualifiers
FT CDS 325..1788
FT /*tag= a
FT /product= stauflen protein
XX CA2238656-A1.
XX PD 22-NOV-1999.
XX
XX 22-MAY-1998; 98CA-2238656.
XX 22-MAY-1998; 98CA-2238656.
XX

PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83024.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1c; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;

Query Match 37.5%; Score 1255; DB 21; Length 2859;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1419; Conservative 0; Mismatches 210; Indels 8; Gaps 3;

QY 454 agaaagcataaccctactgtagaactaaatgcaactgtgcatgaaacttggaaaaaacc 513
Db 264 agaaagtatagcttctcaccattgagctcaatgcactgtgtgtgaacttggaagaaacc 323
QY 514 aatgtataagcctgttgacccttactctcgatgacgtccacctataactacaacatgag 573
Db 324 aatgtataagcccggtgagccctcactctcgatgacgtccacctataactatgcatgcg 383
QY 574 aggaagtgctatcccccgaggtactttaccattccagttccaccttacttatca 633
Db 384 tggaggtgcctatcccccgaggtactttaccattccagttccaccttacttatca 443
QY 634 agtgaacttctgttggaggagacagaacttaatgycaaagaaagacaagacagctgc 693
Db 444 agtgaagctctcgttggcgcgacagcagtttaatgycaaagaaagatgagaccacgct 503
QY 694 gaaacacgatgctgtgccaagcggttgagatcctgacgaatgagccctgcagagag 753
Db 504 gaaacacgatgcccccgccgctgagctgagactctgcagagtgaacccctgcagaaag 563
QY 754 gctggaggtgaatggaagagaatccgaaagagaanaatctcaataatctgaataagtca 813
Db 564 gttggaggtgaatggaagagaagacagagaaagaaacctcaataatcggaaataagcca 623
QY 814 agtgtttgagatgcacttaaacggaacttgctgtgaattcgaggtgcccggagagag 873
Db 624 agtgtttgaatgctgctgaagcgaatctgcctgtgaattttgaggtgcccggagagag 683
QY 874 tggcccaacccacatgaagaactttgtgaccaaggttctcggtggggaggtttgtgggga 933
Db 684 tggcccaacccacatgaagaactttgtgaccaaggttctcggtggggaggtttgtgggga 743
QY 934 aggtgaagggaaagcaagaagatttcaagaanaaatgcccgcataagctgttcttgagga 993
Db 744 aggaagaaggaaagcaagaagatttcccaagaagaatgcccgcagggctgttcttgagga 803

QY 994 gctgaagaagttaccgccccctgctgctgacgttgaaacgagtaaaagcctagaatcaaaaagaa 1053
Db 804 gcttagagagctgcacacccctccctgctgtgagcgagtgaaagcccaagaatcaagaagaa 863
QY 1054 aacaaaaccatagtcgaagccacagacaagccagatatgtgccagggatcaatccgat 1113
Db 864 aagtcagcccacctgca-----agacagccccggattatgtgccaaaggtatgatccat 917
QY 1114 tagccgacttgcccagatccagcagcgaanaagagaagagagccagagttacacgctcct 1173
Db 918 tagtagacttgcaacagatccagcagcgaanaagagaagagagccagagttacatgtcctc 977
QY 1174 cacagagcagagcctcccgccgcccgcagggaggtttgtatgcaggtgaaggttgaaacca 1233
Db 978 tacagaacgaggtcttccacgctgcgaaggaggtttgtatgcaggtgaaggttgggcatca 1037
QY 1234 cactgcagaagaagcggcaccacaagaaggtggccaagcgcaatgcagccgagaacat 1293
Db 1038 cactgcagaagaagtggttaccaataagaagtgcccaagcgtlaatgctgtgagaacat 1097
QY 1294 gctgagatccttgtttcaaaagtcccgccagcggcagcccaacccgactcaagtc 1353
Db 1098 gctgagatccttgggttcaaaagttcccgccagcggcagcctgcgaagccagcactcaatc 1157
QY 1354 agaggaagaagacaccataaagaacccagggatlggaanaagftraaccttttgaacc 1413
Db 1158 agaagaagaagactccagtaaaagaacccaggaagacggaagaagftraacgttttgaacc 1217
QY 1414 tggctctgggagatgaaatgagactagtaataaagagatgagttcaggtgccttatct 1473
Db 1218 tagccctggggatgaaaatggaactagtaacaaggaagcagagagttcaggtgccttatct 1277
QY 1474 aagtcacagcagcctgctgctgtggaattcttcccatggtgcccagagtcgcccagctgt 1533
Db 1278 tagccatcagcagcctgctgctgtggaattcttcccatggtgcccagagtcgcccagctgt 1337
QY 1534 aggaagtagtcaagacatcacaccaagaatttaccaggcagctccgaatcctgcgcaa 1593
Db 1338 cggggttagtcaagacacacacaccaagaatttaccaggcagctccgaatcctgcgcaa 1397
QY 1594 ggcacggtlaactgcatagtatagcccgagagttgtgtatggggcaccctgcgcccacagc 1653
Db 1398 ggcacggtlaactgcatagtatagcccgagagttgtgtatggggcaccctgcgcccacagc 1457
QY 1654 cgagaccattttaagaataacatctcttcaggccaagtlaccocatggaacctctcacgag 1713
Db 1458 cgagaccattttaagaagataacatctcttcaggccaagtlaccocatggaacctctcacgag 1517
QY 1714 accctctgagcaactggaactatcttccagagttccaggatccagggttgaataacaaga 1773
Db 1518 accctctgagcaactggaactatcttccagagttccaggatccagggttgaataacaaga 1577
QY 1774 ctcccccaaaaacaagaagaactttgtatcttatacgaattgtctctccagccacc 1833
Db 1578 ctcccccaaaaacaagaagaactttgtatcttatacgaattgtctctccagccacc 1637
QY 1834 tctgatacagcagtgatcggcaagagtgtagtccctgcacatgatatgctgcgctgaa 1893
Db 1638 tctgatacagcagtgatcggcaagagtgtagtccctgcacatgatatgctgcgctgaa 1697
QY 1894 catcttaaagttgctgtgtgaggttggaacaaagttacagagatgccaagaacagagaa 1953
Db 1698 catcttaaagttgctgtgtgaggttggaacaaagttacagagatgccaagaacagagaa 1757
QY 1954 cggaccaatgtctgtgtgtggaaggtgctgaaacctttctgccaatgaaacctataaaa 2013
Db 1758 tggaccagtttccagcgttcgggaaggtgctgaaacctttctgccaatgaaacctataaaa 1817
QY 2014 tcccaacatatatactgaaataactga-aactgctttgaaatlttgaaattctgtatacc 2072
Db 1818 -cccaacatatatactgaaataactggaactgctttgaaatlttgaaatctgtatatac 1876
QY 2073 tccagtgggccgagaga 2089

Db 1877 tccagtggcccaagaca 1893

|||||

RESULT 7

ID X39991 standard; DNA; 1127 BP.

XX X39991;

DT 02-JUL-1999 (first entry)

XX

DE Prostate cancer associated gene.

XX

KW Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer; ss.

XX

OS Homo sapiens.

XX

PN WO9904265-A2.

XX

PD 28-JAN-1999.

XX

PF 15-JUL-1998; 98WO-US14679.

XX

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ; Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E; Tureci O;

PI

XX

DR WPI; 1999-132448/11.

XX

PT New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products for the diagnosis, monitoring or treatment of cancers

PS Claim 67; Page 625-626; 787pp; English.

XX

XX The invention relates to a method for diagnosing a disorder characterised by expression of a human cancer associated antigen precursor coded for by a nucleic acid molecule (NAM). The method comprises: (a) contacting a biological sample isolated from a subject with an agent that specifically binds to the NAM, an expression product or a fragment of an expression product complexed with an HLA molecule; and (b) determining the interaction between the agent and the NAM or the expression product as a determination of the disorder. The products and methods can be used in the diagnosis, monitoring, research, or treatment of conditions characterised by the expression of various cancer associated antigens. The invention provides nucleic acid sequences and encoded polypeptides which are cancer associated antigen precursors expressed in human breast cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and lung cancer.

XX

SQ Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;

Query Match 33.1%; Score 1107.2; DB 20; Length 1127; Best Local Similarity 99.4%; Pred. No. 7.2e-266; Matches 1120; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 651 gagagacagcaatttaatgcaaaaggaaagacagagctgcgaaacacagatgctgctg 710

Db 1 gagagacagcaatttaatgcaaaaggaaagacagagctgcgaaacacagatgctgctg 60

QY 711 ccaaaagcgttgaggatcctgacgaatgagccctgccaagagagctggaagtgaatgaa 770

Db 61 ccaaaagcgttgaggatcctgacgaatgagccctgccaagagagctggaagtgaatgaa 120

QY 771 gagaatccgaagaagaataatcgaataatcgaataaagtgaagtgttgagattgcac 830

Db 121 gagaatccgaagaagaataatcgaataatcgaataaagtgaagtgttgagattgcac 180

QY 831 ttaacggaacttgctgtgaatttcgaggtggccgggagagtggtgcccaacatga 890

Db 181 ttaacggaacttgctgtgaatttcgaggtggccgggagagtggtgcccaacatga 240

QY 891 agaacttgtgaccaaggttctcggttggtggaggttggtgggaagtggaaggaaagca 950

Db 241 agaacttgtgaccaaggttctcggttggtggaggttggtgggaagtggaaggaaagca 300

QY 951 agaagattcaaaagaaaatgcccgcataagctgttcttgaggagctggaagaattaccgc 1010

Db 301 agaagattcaaaagaaaatgcccgcataagctgttcttgaggagctggaagaattaccgc 360

QY 1011 ccctgctgcagttgaaacagtaaaagcctagaaatcaaaaagaaaacaaacccatagtca 1070

Db 361 ccctgctgcagttgaaacagtaaaagcctagaaatcaaaaagaaaacaaacccatagtca 420

QY 1071 agccaacagacaagcccagaatatgcccaggggatccaatccgattagccgactggcccaga 1130

Db 421 agccaacagacaagcccagaatatgcccaggggatccaatccgattagccgactggcccaga 480

QY 1131 tccagcaggcacaaaagagagagagagcgcagagtaacacgctcctcacagagcgagccctcc 1190

Db 481 tccagcaggcacaaaagagagagagagcgcagagtaacacgctcctcacagagcgagccctcc 540

QY 1191 cgcgcgcgagggaggttgatgacagtgtaaggttgtaaaaccacactgcagaagaagcgg 1250

Db 541 cgcgcgcgagggaggttgatgacagtgtaaggttgtaaaaccacactgcagaagaagcgg 600

QY 1251 gcaccaacaagaaggtgcccgaagcgcaatgcaagccgaagaacatgctgagatccttggt 1310

Db 601 gcaccaacaagaaggtgcccgaagcgcaatgcaagccgaagaacatgctgagatccttggt 660

QY 1311 tcaaaagtccgcagcgcgcacccaacccgcgactcaagtcagagagagaagacacca 1370

Db 661 tcaaaagtccgcagcgcgcacccaacccgcgactcaagtcagagagagaagacacca 720

QY 1371 taagaacaaccaggggatggaagaagaataacctttttgaacctgctctggtggatgaa 1430

Db 721 taagaacaaccaggggatggaagaagaataacctttttgaacctgctctggtggatgaa 780

QY 1431 atggaactagtaataaagagatgagttcagagatgccttatcgaatcatcaagcagctgc 1490

Db 781 atggaactagtaataaagagatgagttcagagatgccttatcgaatcatcaagcagctgc 840

QY 1491 ctgctggaattcttcccatggtgcccagagtcgcccaggtctgtagagttgaagagac 1550

Db 841 ctgctggaattcttcccatggtgcccagagtcgcccaggtctgtagagttgaagagac 900

QY 1551 atcacaccaagaattttaccaggggcagctccgaatcctgccaagggccaagcgttaactgcca 1610

Db 901 atcacaccaagaattttaccaggggcagctccgaatcctgccaagggccaagcgttaactgcca 960

QY 1611 tgatagcccgagagtggtgtgtagtgggggcacctgcgccacag-ccgagaccattttaag 1669

Db 961 tgatagcccgagagtggtgtgtagtgggggcacctgcgccacagcccgagaccattttaag 1020

QY 1670 aataacatctctcagggccagtaaccccatggaacctcacaagagaccctctgaagcaactg 1729

Db 1021 aataacatctctcagggccagtaaccccatggaacctcacaagagaccctctgaagcaactg 1080

QY 1730 gactatcttccagaggtccagggattccaggttgaaatacaagaactt 1776

Db 1081 gactatcttncagaggtccagggattncaggttgaaatacaagaactt 1127

RESULT 8
X40002
ID X40002 standard; DNA; 946 BP.
XX
AC X40002;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahn U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629-630; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other;

Query Match 23.9%; Score 798.8; DB 20; Length 946;
Best Local Similarity 93.6%; Pred. No. 3.7e-189;
Matches 888; Conservative 0; Mismatches 54; Indels 7; Gaps 6;

QY 651 gagacagcaatttaatgcaagaagacaagacagctgcgaaacacgatgctg 710
|||||
DB 1 gagacagcaatttaatgcaagaagacaagacagctgcgaaacacgatgctg 60
QY 711 ccaagcgttgagatcctgcagaatagccctgcagagagctgagtgtaatgaa 770
|||||
DB 61 ccaagcgttgagatcctgcagaatagccctgcagagagctgagtgtaatgaa 120

QY 771 gagaatccgaagaagaaaatcacaataaactgaataaagtcagtgcttgagatgac 830
|||||
DB 121 gagaatccgaagaagaaaatcacaataaactgaataaagtcagtgcttgagatgac 180
QY 831 ttaacggaacttgctgtgaatttcgaaggtgccccggagagtgccccaccacatga 890
|||||
DB 181 ttaacggaacttgctgtgaatttcgaaggtgccccggagagtgccccaccacatga 240
QY 891 agaacttgtaaccaaggttcggttgaggagttgttgagggaaggtgaagggaagaagca 950
|||||
DB 241 agaacttgtaaccaaggttcggttgaggagttgttgagggaaggtgaagggaagaagca 300
QY 951 agaagattcaagaagaaatgcccccatagctgttcttgaggagctgaagaagttaccgc 1010
|||||
DB 301 agaagattcaagaagaaatgcccccatagctgttcttgaggagctgaagaagttaccgc 360
QY 1011 ccctgctgcagttgaacgagtaagcctagaatcaaaaagaaaacaaccatagtca 1070
|||||
DB 361 ccctgctgcagttgaacgagtaagcctagaatcaaaaagaaaacaaccatagtca 420
QY 1071 agccacagacaagcccaagaatatgcccagggtatcaatccgattagccgactgcccaga 1130
|||||
DB 421 agccacagacaagcccaagaatatgcccagggtatcaatccgactgcccaga 480
QY 1131 tccagcagcaaaaaagagagaagagccagagtlacacgctcctcaacagagcgagcctcc 1190
|||||
DB 481 tccagcagcaaaaaagagagaagagccagagtlacacgctcctcaacagagcgagcctnc 540
QY 1191 cgcgcgcagagagttgtgatgcaggtgaaggttggaaccacac-tgcagaagagaaacg 1249
|||||
DB 541 cgcgcgcagagagttgtgatgcaggtgaaggttggaaccacac-tgcagaagagaaacg 600
QY 1250 ggcaccacaagaaggtgycacaagcgcaatgcaagcagagaacatgctgagatccttgt 1309
|||||
DB 601 ggcaccacaagaaggtgycacaagcgcaatgcaagcagagaacatgctgagatccttgt 660
QY 1310 ttc-aaagtcctgcagcgcaagcccaacacccgcactcaagtcagaggaagaacaccc 1368
|||||
DB 661 ttcaaaagtcctgcagcgcaagcccaacacccgcactcaagtcagaggaagaacaccc 720
QY 1369 cataaagaacacagggatggaagaagaatlaaccttttgaacctgctcgtgagatga 1428
|||||
DB 721 altaagaacacanggatggaagaagaatlanc-ntttgaanctgctnltgagatta 778
QY 1429 aaatggaactgaataaagaagatgagttcagatgcttatctaatcatcagcagct 1488
|||||
DB 779 aaatgg-cctgtantaagaagatgagttcagatgctnctatctaagtcataaaca-cl 836
QY 1489 gcctgctgaatcttcocatggtgcccagagtgcccagcgtgtagaggttagtaag 1548
|||||
DB 837 tgctgctgaatcttcocatggtgcccagagtgcccagcgtttagaggtla-tcang 895
QY 1549 acatcacaccaagaatttaccagggcagctccgaatcctgccaagggc 1597
|||||
DB 896 ccnttcnccaagaatttcccgggcaggttccaatctgccaagggc 944

RESULT 9
X40000
ID X40000 standard; DNA; 844 BP.
XX
AC X40000;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM prostate cancer; ss.
OS Homo sapiens.
XX

PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

Query Match 22.6%; Score 755.6; DB 20; Length 844;
Best Local Similarity 96.3%; Pred. No. 1.9e-178;
Matches 812; Conservative 0; Mismatches 26; Indels 5; Gaps 4;

QY 813 aagtgttgatgtgcacttaacggaactgctgtgaatttcgaggtgcccgggaga 872
|||
Db 1 aagtgttgatgtgcacttaacggaactgctgtgaatttcgaggtgcccgggaga 60

QY 873 gtggccacccccacatgaagaactttgtgaccaagtttcggtgggagtttgtgggg 932
|||
Db 61 gtggccacccccacatgaagaactttgtgaccaagtttcggtgggagtttgtgggg 120

QY 933 aagtgaaaggaaaagcaagaatttcaaaagaaaatgccgcatagtctgtctgagg 992
|||
Db 121 aagtgaaaggaaaagcaagaatttcaaaagaaaatgccgcatagtctgtctgagg 180

QY 993 agctgaagaagttaacgcgcctctgcctgtgaacgagtaaaagcctagaatcaaaaaga 1052
|||
Db 181 agctgaagaagttaacgcgcctctgcctgtgaacgagtaaaagcctagaatcaaaaaga 240

QY 1053 aaacaaaccatagtcaagccacagacagccacagatatgtgccaggggatcaatccga 1112
|||
Db 241 aaacaaaccatagtcaagccacagacagccacagatatgtgccaggggatcaatccga 300

QY 1113 ttagccgactgtgccagatccagcagcgaataaaaggagaagagccagagtacacgctcc 1172
|||
Db 301 ttagccgactgtgccagatccagcagcgaataaaaggagaagagccagagtacacgctcc 360

QY 1173 tcacagagcgagcctcccgccgccgcaaggagtttgtatgcaggtgaagtttgaaac 1232
|||
Db 361 tcacagagcgagcctcccgccgccgcaaggagtttgtatgcaggtgaagtttgaaac 420

QY 1233 acaactgcaagaagcagcgccaccaagaaggctggccaagcgcaatgcagccgagaaca 1292
|||
Db 421 acaactgcaagaagcagcgccaccaagaaggctggccaagcgcaatgcagccgagaaca 480

QY 1293 tgcctgagatcccttggttccaagtcccgccgagcgccaccaaccgcactcaagt 1352
|||
Db 481 tgcctgagatcccttggttccaagtcccgccgagcgccaccaaccgcactcaagt 540

QY 1353 cagaggaagaagacaccataaagaacccagggatggaagaaaagtaacctttttgaac 1412
|||
Db 541 cagaggaagaagacaccataaagaacccagggatggaagaaaagtaacctttttgaac 600

QY 1413 ctggccttggggatgaaatgggactagtaataaagagatgagttcaaggatgcctatc 1472
|||
Db 601 ctgctcttgggatgaaatgggactagtaataaagagatgagttcaaggatgcctatc 660

QY 1473 taagtcatacagcagctgcctgtggaattcttcccatggtgccgaggtgccagctg 1532
|||
Db 661 taagtcatacagcagctgcctgtggaattcttcccatggtgccgaggtgccagctg 720

QY 1533 taag-agttagtcaagacatcacaccaagaatttaccaggcgagtcgccaatcctgcc 1591
|||
Db 721 taggaagttagtcaagacatnacaccaagaatttaccag-gccagcttcgaatcctgcc 779

QY 1592 aagccacggtactgcattgatagcccgagagtttgttatggggcaactcgcacaca 1651
|||
Db 780 anggcncng--tactgcattgatagccc-anagtttgttatggggcaanccttgcccca 836

QY 1652 gcc 1654
|||
Db 837 ggc 839

RESULT 10
X39992
ID X39992 standard; DNA; 773 BP.
AC X39992;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX
PS Claim 67; Page 626; 787pp; English.
CC
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

Query Match 22.5%; Score 753.2; DB 20; Length 773;
Best Local Similarity 99.1%; Pred. No. 7.3e-178;
Matches 766; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1005 taccgccccctgctgagttgaacagtaaaagcctagatacaaaaagaacaaccaca 1064
DB 1 taccgccccctgctgagttgaacagtaaaagcctagatacaaaaagaacaaccaca 60
QY 1065 tagtcaagccacagacaagcccaagataatgcccagggatcaatccgattagccgactgg 1124
DB 61 tagtcaagccacagacaagcccaagataatgcccagggatcaatccgattagccgactgg 120
QY 1125 ccagatccagcagcagcaaaaagagagagcagagatcacgctcctcacagagcgag 1184
DB 121 ccagatccagcagcagcaaaaagagagagcagagatcacgctcctcacagagcgag 180
QY 1185 gcctcccgccgcccagggagttgtgatgcaggtgaaggttggaaccacactgcagaag 1244
DB 181 gcctcccgccgcccagggagttgtgatgcaggtgaaggttggaaccacactgcagaag 240
QY 1245 gaacgggcaccaacaagaaggtgcccagcgaatgcagccgagaacatgctgagatcc 1304
DB 241 gaacgggcaccaacaagaaggtgcccagcgaatgcagccgagaacatgctgagatcc 300
QY 1305 ttggttcaaaagtcgcgcagcgcacccaaccgccactcaagtccagagagaaga 1364
DB 301 ttggttcaaaagtcgcgcagcgcacccaaccgccactcaagtccagagagaaga 360
QY 1365 caccataaagaacacaggggatggaagaagaagtaaccttttgaacctgctctgggg 1424
DB 361 caccataaagaacacaggggatggaagaagaagtaaccttttgaacctgctctgggg 420
QY 1425 atgaaaatgggactagtataaagaagatgagttcagatgccttatctaagtcacagc 1484
DB 421 atgaaaatgggactagtataaagaagatgagttcagatgccttatctaagtcacagc 480
QY 1485 agctgcctgctgtaattcttccatggtgccgaggtgcgccagggctgtagagttagtc 1544
DB 481 agctgcctgctgtaattcttccatggtgccgaggtgcgccagggctgtagagttagtc 540
QY 1545 aaggacatcacaccaagaatttaccagggcagctccgaatcctgccaagggcaggttaa 1604
DB 541 aaggacatcacaccaagaatttaccagggcagctccgaatcctgccaagggcaggttaa 600
QY 1605 ctgcatgatagcccgagagttgtgttatggyggcacctgcgccacag-ccgagaccatt 1663
DB 601 ctgcatgatagcccgagagttgtgttatggyggcacctgcgccacagcccgagaccatt 660

QY 1664 ttaaagaataacatctcttcaggccacgtaccatggacctctcacagacccttgag 1723
DB 661 ttaaagaataacatctcttcaggccacgtaccatggacctctcacagaccctntgag 720
QY 1724 caactgactatacttccagaggtccaggtatccaggttgaatacaagaact 1776
DB 721 caactgactatacttccagaggtccaggtatccaggttgaatacaagaact 773

RESULT 11
X40001/c
ID X40001 standard; DNA; 825 BP.
XX
AC X40001;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX
PS Claim 67; Page 629; 787pp; English.

XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 825 BP; 289 A; 152 C; 151 G; 228 T; 5 other;

Query Match 16.5%; Score 551; DB 20; Length 825;
Best Local Similarity 95.9%; Pred. No. 1.6e-127;
Matches 585; Conservative 0; Mismatches 23; Indels 2; Gaps 2;

Accession	Species	Gene	Position	Sequence	Length
QY 2731	ttcttattcattgactaaccacagacaglttcagtgatgcgaattgltgcccctcgt		2790		
Db 791	TTTAATTTCCTTGGCTAAACCCAGCAGNCAGTTCAGTGATGCAATTGGGGCCNCTGGT		732		
QY 2791	tcagctgaaac-cagtcctggaacttcaaaaacctggaataagctccacagltgtataa		2849		
Db 731	TCAGCTGAACCAAGTCCCTGAGCTTCAAAAACCTTGAATAAGTCTCCCCAGTTGTATAA		672		
QY 2850	attgacaacatttaggaattttaacatttagatgatcatttggttccatttttatttcatt		2909		
Db 671	ATTGAGACCATTTAGGAATTTTAAACITTAGATGATCATTTGGTCCATTTTTATTNTCATTT		612		
QY 2910	tttatttttgttaaatgcaaacagagc-ttaaatagaactttgatctcgttlttaagaatta		2968		
Db 611	TTTATTTTTGGTTAATGCAAAACAGAGACTTAAATGAACCTTGANCTCTGTTTTAAAGATTA		552		
QY 2969	ttaaaaaacatgtgtatctatatacatatgagctccttgagagacttagctttcactacactac		3028		
Db 551	TTAAAAAACATTTGTGTATCTATACATATGGCTCTTGAGGACTTAGCTTTCACACTACACTAC		492		
QY 3029	aggatatgatctccatagttagtccataataaacctgcagagatgattttccagagatgctgat		3088		
Db 491	AGGATATGATCTCCATGTAGTCCATATTAACCTGCAGAGTGAATTTTCCAGAGTGCTCGAT		432		
QY 3089	actgtlaattacatctccattagagctgaaagaatgacctaagttctgtatacagctg		3148		
Db 431	ACTGTAAATACATCTCCATTAGGGCTGAAAGAATGACCTACGTTTCTGTATACAGCTG		372		
QY 3149	tgttgctttgatgtgtgttactgtacacagaagtgtgtgcactgaggtcctgctgtgtg		3208		
Db 371	TGTTGCTTTGTGATGTGTGTACTGTACATGACAGAAAGTGTGTGCACTGAGGCTCTGCGTGTG		312		
QY 3209	gtccgtatgaaaaacctggtagccctgcgaggttaagtactgtctccattcattgtttacg		3268		
Db 311	GTCGATATGGAAGCCCTGGTAGCCCTGCAGAGTTAAGTACTGCTTCATTCATTGTTTACG		252		
QY 3269	ctggaattttctcccccattggaatgttaagtaaaacttaagtgtttgtcatcaataaatg		3328		
Db 251	CTGGAATTTTCTCTCCCATGGAATGTAAGTAAAACTTAAGTGTGTCATCAATAAATGG		192		
QY 3329	taataactaaa 3338				
Db 191	TAATACTAAA 182				
RESULT 12					
ID X39993/c					
XX X39993 standard; DNA; 794 BP.					
AC X39993;					
XX 02-JUL-1999 (first entry)					
DT Prostate cancer associated gene.					
DE Cancer associated antigen; diagnosis; research; treatment; human;					
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;					
KW prostate cancer; ss.					
XX Homo sapiens.					
OS WO9904265-A2.					
XX 28-JAN-1999.					
PD 15-JUL-1998; 98MO-US14679.					
XX 22-JUN-1998; 98US-0102322.					
PR 17-JUL-1997; 97US-0896164.					
PR 10-OCT-1997; 97US-0061599.					
PR 10-OCT-1997; 97US-0061765.					
PR 10-OCT-1997; 97US-0948705.					

```

PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
PA
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,
PI pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI; 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67; Page 626; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 794 BP; 273 A; 141 C; 152 G; 206 T; 22 other;
XX
Query Match 15.0%; Score 501.4; DB 20; Length 794;
Best Local Similarity 96.3%; Pred. No. 3.4e-115;
Matches 517; Conservative 0; Mismatches 18; Indels 2; Gaps 1;
QY 2804 tctctgacttccaacaccttgaatgaagtcctcc--acagtgtataaatgacattt 2861
DB 718 TCCTGGACTTTCAAAACCCCTGGAATGAAGTCTCCCCNCCAGTTGNAATAATTGGCCNATTT 659
QY 2862 aggaattttaaactltagatgatcatttggtlccatttllattcatttattt 2921
DB 658 AGGAATTTTAACNTTAGATGTTCATTGGTTCATTTTATTTTATTTTGGT 599
QY 2922 aatgcaaacagagacttaaatgaactttgatctctgttlltaagaattataaaacattg 2981
DB 598 AATGCAACACAGGNCCTTAATGAACCTTTCATCTCTGTTTAAAGATTATTAACCAATGTG 539
QY 2982 tgtatctatacatatggtcctctgaagacttagcttccactacactacagatatgctc 3041
DB 538 TGNATCTATACATATAGGCTCTTGAGAGACTTAGCTTTCACCTACACTACAGGATATGATCTC 479
QY 3042 catgtagtccatataaaccctgcagagtgaatttccagagtgcctgatactgttaattaca 3101
DB 478 CAGTAGTCCATATATAAACCTGCAGAGTGAATTTTCCANAGTGCCTGATACTGTTAATTACA 419
QY 3102 tctccattagggctgaagaatgacactacgcttctgtatatcacgctgtgttcttgat 3161
DB 418 TCTCCNTTAGGGCTGAAAAGAATGACCTACGTTCTTGATATACANCTGTGTGCTTTGAT 359
QY 3162 gttgtgttactgttacacagaagtgtgtgcactgaggtcctgcgtgtgtgtccgtatgaaa 3221
DB 358 GTTGCTGNTACTGNACACAGAGGTGTGTGCNCTGAGGCTCTGCGGTGTGTCCTGGAATA 299
QY 3222 acctgttagccctgcgagtgtaagtactgtcttccattcattgtttaaagcttgaatttct 3281
DB 298 GCCTGTAGCCCTGCGAGTTAAGTACTGCTTCATTCATGTGTTTACGCTGGAATTTTCT 239
QY 3282 ccccatggaatgtaagtaaaacttaagtgttgtcatcaataaatgylaactaaa 3338
DB 238 CCCCATGGAATGTAAGTAAACCTTAAGTGTGTGCATCAATNAATGTAATACTAAA 182

```


RESULT	13
279981/c	
ID	279981 standard; CDNA; 660 BP.
XX	
AC	279981;
XX	
DT	07-APR-2000 (first entry)
XX	
DE	Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX	
KW	Human; gene expression product; diagnosis; tumour; colon cancer;
KW	colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW	hyperplasia;.ds.
XX	
OS	Homo sapiens.
XX	
PN	WO9964576-A2.
XX	
PD	16-DEC-1999.
XX	
PF	09-JUN-1999; 99WO-IB01062.
XX	
PR	10-JUN-1998; 98US-0088801.
XX	
PA	(FARB) BAYER CORP.
XX	
PI	Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI	Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI	Schlegel R;
XX	
DR	WPI; 2000-087220/07.

PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -
XX
XX
PS Claim 1; Page 158; 469pp; English.
XX
CC 279917 to Z80766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other;

Query Match	13.0%;	Score 433.8;	DB 21;	Length 660;
Best Local Similarity	92.4%;	Pred. No. 2.1e-98;		
Matches 495; Conservative	0;	Mismatches 36;	Indels 5;	Gaps 4;

OY	1157	ccagagtacacgcctctctcaacagacgagcgctcccgccgcgagggagtttgtatgcag	1216
Db	531	CCAAGGTACACGTTCTTCANGAAGCNAAGCCTTCGGNGCGNAGAAGATTGTATATNCAG	472
OY	1217	gtgaaggttgnaaccacactgcagaaggaacgycacccaagaagtgtgcacacgc	1276
Db	471	GTNAAGGTAGAACCCCATTTGCANAAAGGA--CGGGCACCCACAAGAAGGTGGCCAAG-GC	415
OY	1277	aatgcagccgagacaatgctgagatccttgytttcaaagtccgcgacgycgacccacc	1336
Db	414	AATGCAG-CGAGANCATGTTGAAGATCCTGGGTTTCAAAAGTCCGCAGGGCGCAG-CCACC	357
OY	1337	aaaccgcactcaagtcagagagagaacacaccataaagaacacagggatgtgaagaaaa	1396

Db	356	AAACCGGCACTCAAGTCAAGAGAGAGAAGACACCATTAAAGAAACCAGGGGATGGAAGAAAA	297
QY	1397	gtaaccttlttgaacctgctctgggatgaaatggaactagtaataagagatgag	1456
Db	296	GTAACCTTTTGAACCTGGCTCTGGGGATGAAATGGGACTAGTAATAAAGAGAGTGA	237
QY	1457	ttcagatgccttatctlaagtcatacagcagctgcctgctggaattcttcccatgtgccc	1516
Db	236	TTCAGGATGCCTTATCTAAGTATCAGCAGCTGCCTGCTGGAATTCTTCCCATGGTGCCC	177
QY	1517	gaagtcgcccagcgtgttagagattagtcgaagacatcacaccaagaatttaccagggca	1576
Db	176	GAGGTGCCCCAGGCTGTAGGAGTTAGTCAAGGACATCACACCAAGATTTTACCAGGGCA	117
QY	1577	gtcccgatcctctgccaagggccaacgctaactgcatgatagcccagagagtgtgtatggg	1636
Db	116	GCTCCGAATCCTCGCCAAAGGCCACGGTAAC TGCCATGATAGCCCGAGAGTGTGTATGGG	57
QY	1637	ggcacctcgcccacacagccggaacattttaagaataacatctcttcagggccagct	1692
Db	56	GGCACCTCGCCACACAGCCGAGACCAATTTTAAAGATAAACA TCTCTTCAGGGCCACGT	1

RESULT	14
Z80269	
ID	Z80269 standard; cDNA; 727 BP.
XX	
AC	Z80269;

DT	07-APR-2000	(first entry)
XX		
DE	Human colon cancer cell line SW480	cdna clone SEQ ID NO:353.

KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.

OS Homo sapiens.

PN W09964576-A2.

PD 16-DEC-1999.

09-JUN-1999; 99WO-IB01062.

PR 10-JUN-1998; 98US-0088801.

PA (FARB) BAYER CORP.

PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE,
PI Schlegel R;

WPI; 2000-087220/07.

Novel nucleic acids, used to develop products for the diagnosis and treatment of disorders involving unwanted cell proliferation, particularly cancers, especially colon cancer

Claim 15; Page 273; 469pp; English..

CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.

SQ Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;

Query Match 12.3%; Score 412.2; DB 21; Length 727;
Best Local Similarity 95.5%; Pred. No. 5.1e-93;
Matches 444; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 594 ggtacttttaccattccagttccaccttacttatacagtggaacttctgtggag 653
Db 1 ggtacttttaccattccagttccaccttacttatacagtggaacttctgtggag 60
QY 654 gacagcaatttaatgcaaaagaaagacaagacagctgcgaacacagatgctgccca 713
Db 61 gacagcaatttaatgcaaaagaaagacaagacagctgcgaacacagatgctgccca 120
QY 714 aagcgttgagatcctgcgaatgagccctgcgacagagagctggagtgatgaaagag 773
Db 121 aagcgttgagatcctgcgaatgagccctgcgacagagagctggagtgatgaaagag 180
QY 774 aatccgaagaagaatatctcaataatctgaataagtcagtgtttgagattgcactta 833
Db 181 aatccgaagaagaatatctcaataatctgaataagtcagtgtttgagattgcactta 240
QY 834 aacggaacttgctgtgaatttcgaggtgcccgggagagtgccccacccacatgaaga 893
Db 241 aacggaacttgctgtgaatttcgaggtgcccgggagagtgccccacccacatgaaga 300
QY 894 actttgtgaccaaaggttcggttgaggagttgtgggggaaagtggaaggaaaaagcaaga 953
Db 301 actttgtgaccaaaggttcggttgaggagttgtgggggaaagtggaaggaaaaagcaaga 360
QY 954 agatttcaaaagaaatg-cgcgcataagctgttcttgaggagctgaagaagtaccggccc 1012
Db 361 agatttcaaaagaaatgcccgcataagctgttcttgaggagctgaagaagtaccggccc 420
QY 1013 ctgcctgcagltgaac---gagtaaagcctagaatcaaaaaagaaa 1054
Db 421 ctgcctgnattgacgcgaagttaagccctanaatccaatgaaa 465

RESULT 15
280703/c
ID 280703 standard; cDNA; 708 BP.

AC 280703;
DT 07-APR-2000 (first entry)

DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:787.

KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytosolic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.

OS Homo sapiens.

PN WO964576-A2.

PD 16-DEC-1999.

PF 09-JUN-1999; 99WO-IB01062.

PR 10-JUN-1998; 98US-0088801.

PA (FARB) BAYER CORP.

PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;

WPI; 2000-087220/07.

PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer

PS Claim 15; Page 444; 469pp; English.

CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.

SQ Sequence 708 BP; 226 A; 127 C; 143 G; 191 T; 21 other;

Query Match 9.8%; Score 327.8; DB 21; Length 708;
Best Local Similarity 82.8%; Pred. No. 5.1e-72;
Matches 515; Conservative 0; Mismatches 83; Indels 24; Gaps 12;

QY 2577 tttctttgattgatatgacac--tatataaatcttcattgagaattctc---aat 2631
Db 622 TTTCCNTTNGATTGAATGGCCCCATAATAAAATTTCCNTTTGAAATTTTCCCAATG 563
QY 2632 tgtatctagttaatatgacacagtttgaaact--gtctgagactgacttacc--aat 2686
Db 562 GGATCCAGTTTAATAGCCCCAGTTGNAACCTTGTTAAAGCCTTGCTTACCAATA 503
QY 2687 aatctaaccgac--aaagatcataatccatgtgtatgtttagacattttattcatgt 2744
Db 502 ATTCTNAACCGGCCAAAGTTCAAAACCNAAGGGGAAGGNGNTAAGCNATTTAATTCATG 443
QY 2745 actaacc--aggacagtttcagtgatgcaaatgtgtccctct-ggttcagctgaaa- 2800
Db 442 NCTTACCCAGGGCCAGTTCAGTTAAGCAAAATTTGGGGCTTCTGGGTTACGCTGAAC 383
QY 2801 cagtcctgactttcaaaaaccttgaa-taagtcctcca-cagttgtataaatgagcaa 2858
Db 382 CAGTCCTGCAGCTTCCAAAAACCCGTAATTAAAGTCTCCACCAAGTTGATAAATGNCNCA 323
QY 2859 tttaggaattttaacttttagatgataca----tttgttccatttttatttcaattta 2913
Db 322 TTTAGGAATTTTAAACCTTTTAGATGATCCATTTGGTTCCCATTTTATTTTAA 263
QY 2914 tttttgttaatgcaaacagagacttaaatgaactttgatctctgttttaagaattataa 2973
Db 262 TTTTGTTAATGCAAAACAGACTTAAATGAACCTTTGATCTCTGTTTAAAGATTAATTAA 203
QY 2974 aaacatgtatatacatatagctctcttgaggacttagcttcaactacacagat 3033
Db 202 AAACATTTGTATCTATACATATGCTCTTGAGGACTTAGCTTCACTACACTACAGAT 143
QY 3034 atgatcctcatgtagtcataataaacctgcagagtgatgtttccagagtgctcgatactgt 3093
Db 142 ATGATCTCCATGTAGTCCATATAAACCCTGCAGAGTGATTTCCAGAGTGCTGATACTGT 83
QY 3094 taattacatcctcattagggctgaaaagaatgacactacggtttctgtatacag-ctgtgtt 3152
Db 82 TAATTACATCTCCATTAGGGCTGAAGAATGACCCACGTTTCTGTATACAGCCTGTGT 23
QY 3153 gcttttgatgttgttactgt 3174
Db 22 GCTTTGATGTGTGTACTGT 1

Search completed: April 5, 2001, 08:32:35
Job time: 59399 sec

US-08-458-568A-11/c
Sequence 11, Application US/08458568A
Patent No. 5821339
GENERAL INFORMATION:
APPLICANT: Schaffer, Priscilla A.
APPLICANT: Yeh, Lily
TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
TITLE OF INVENTION: Infections
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,568A
FILING DATE: 02-JUNE-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/065,146
FILING DATE: 05-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn R.
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: DFCI-0029
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 12001 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Herpes simplex virus
STRAIN: Herpes Simplex Virus Type 1
US-08-458-568A-11

```

Query Match          1.3%; Score 44.2; DB 1; Length 12001;
Best Local Similarity 58.9%; Pred. No. 0.095;
Matches 76; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 90 cccccccccgcgcgcgcgcgcgcctcctccacgcacatccgcctctccctccc 149
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7697 CCCCCCTCCTCGCCCCCGCGCTCCCCCTCCTCCTCGCCCCCGCGTCCCCCTCGGCC 7638

QY 150 ttgcctccctctcctcctcctccttttctcctctcctcctccctcctgcgcacagccc 209
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7637 CGCGTCCCCCTCCTCGCCCCCGCGTCCCCCTCCTCCTCGCCCCCGCGTCCCCCTCCT 7578

QY 210 agcaccgcc 218
      | | | | |
Db 7577 CCGCCACAC 7569

RESULT 12
US-08-323-443B-1
; Sequence 1, Application US/08323443B
; Patent No. 5654170
; GENERAL INFORMATION:

```

APPLICANT: KLINGER, KATHERINE W.
 APPLICANT: LANDES, GREGORY M.
 APPLICANT: BURN, TIMOTHY C.
 APPLICANT: CONNORS, TIMOTHY D.
 APPLICANT: DACKOWSKI, WILLIAM R.
 APPLICANT: GERMINO, GREGORY
 APPLICANT: QIAN, FENG
 TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Darby & Darby PC
 STREET: 805 Third Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/323,443B
 FILING DATE: 12-OCT-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Ludwig, S. Peter
 REGISTRATION NUMBER: 25,351
 REFERENCE/DOCKET NUMBER: 0372/0A462
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 527-7700
 TELEFAX: (212) 753-6237
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 31571 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: PKD1 GENOMIC

```

Query Match      1.28; Score 41.4; DB 1; Length 31571;
Best Local Similarity 53.1%; Pred. No. 1;
Matches 111; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY      1  acttcctgcgagctgcggcgccctgagcgcctctcagcgtttgcgcgcgcgcgcgcgt 60
      ||| ||| | ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      9941 ACTCCCCGAGGGGCTGAGGGGGCCCT--GCGCCAGGTGCAGCTGTGGGTGCTGCCAGC 9998

QY      61  ctctctcgagctccgcgtctcctttgacgcgcctccccccccgcgcgcgcgcgcgcgcctc 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      9999 CCCTTCCACACCTCTCCCTCCCTCCGCGCAGCCCCCTCCACACCTCTCCCTGCCAGCCCCCTCC 10058

QY      121  ctccacgcgcactcgcgcctctctccctcccttcgctccctctctcctctccttttccct 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      10059 CAGCTCTCCCTCCCTCCGCGCAGCCCCCTCCACACCTCTCCCTCCGCGCAGCCCCCTCCACACCTC 10118

QY      181  ctctcctccctcctcctcgcgcgcacacgcgcc 209
      | | | | | ||||| |||
Db      10119 TCCCTCCCTGCGCAGCCCCCTCCACACCTCTC 10147

RESULT 13
US-08-781-891-79
; Sequence 79, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
;
```

APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-79

Query Match 1.2%; Score 41; DB 3; Length 87350;
Best Local Similarity 56.2%; Pred. No. 2.5;
Matches 77; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 62 tctctcgctcccgcttcctttgaccgctcccccccgccgagcgccgctcc 121
DB 55222 TCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 55281
QY 122 tccacggcactccgcttccttccttccttcgcttccttccttccttccttc 181
DB 55282 TTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 55341
QY 182 ttccttccttccttccttcgc 198
DB 55342 TTCTTCTCTCTCTCTC 55358

RESULT 14
US-08-149-096A-1/c
Sequence 1, Application US/08149096A
Patent No. 5556956
GENERAL INFORMATION:
APPLICANT: ROY, A.K.
APPLICANT: CHATTERJEE, B.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO THE
TITLE OF INVENTION: ANDROGEN RECEPTOR GENE AND USES THEREOF
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA

ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,096A
FILING DATE: No. 5556956ember 4, 1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Denise L. Mayfield
REGISTRATION NUMBER: 33,732
REFERENCE/DOCKET NUMBER: UTSK:199\MAY
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2775 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: oligonucleotide
US-08-149-096A-1

Query Match 1.2%; Score 40.8; DB 1; Length 2775;
Best Local Similarity 47.6%; Pred. No. 0.33;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 108 cggcgccgctcctccacagcactccgcttccttccttccttccttccttc 167
DB 2606 CCGCCCCCGCGGCTCCCTGCGCTTCCACACTCTTCTCTCTCCCTCCCTC 2547
QY 168 cccttttcttccttccttccttccttccttcgcccacgagcagcagcagc 227
DB 2546 CCCTCCCTCCCTTCTCTCTTCTTCTCCCTCCCTCACCAGCAAGTAGCGG 2487
QY 228 acgagctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 287
DB 2486 TAGAGACATAGTGTGCGGTGATGTTGAAGGCTGGAAGGACTCCCTGTAG 2427
QY 288 acaacattgttcctggaagccttcctttaaagaagtagaacttagactcata 347
DB 2426 AAGACCCCTTTCCCAATACCACTGTTTTCGGAACCGGCTTAGACTCAAG 2367
QY 348 ctgaattaacct 359
DB 2366 GAAGCTTAACT 2355

RESULT 15
US-08-147-777-3
Sequence 3, Application US/08147777
Patent No. 5914265
GENERAL INFORMATION:
APPLICANT: ROOP, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
: SOFTWARE: WordPerfect (Version 5.1)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/147,777
: FILING DATE:
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: PRIOR APPLICATION DATA: including application
: PRIOR APPLICATION DATA: described below:
: APPLICATION NUMBER: 07/876,289
: FILING DATE: April 30, 1992
: APPLICATION NUMBER: Unassigned (204/144)
: FILING DATE: October 29, 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 204/153
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 24979 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
US-08-147-777-3

```

	Query Match	1.2%;	Score 40.8;	DB 2;	Length 24979;
	Best Local Similarity	59.7%;	Pred. No. 1.3;		
	Matches	86; Conservative	0; Mismatches	57; Indels	1; Gaps
QY	54	tgcgagtcctcctcgagctccgacttccatttgacgcgactcccccccggcccgcgcgcg	113		
Db	6957	TGCTCCTTGCGCTCCCTCCGGGCTTCCTCCTCCTCCCTCCGGGCTTCCTCCTCCGCCGC	7016		
QY	114	cgcgctcctcc-acggccactcgcgctctccacctccacctgctccctttctcctccctt	172		
Db	7017	GCTCCTCCCTCTTCTTCTTCTTCTCCTTCTCCTCTCATTTCTTCATCTTCATCTTCATC	7076		
QY	173	tttccctctctcccttccccctcctc	196		
Db	7077	TTCTCCTCTTTTCTTCTTCTTC	7100		

Search completed: April 5, 2001, 09:06:15
Job time: 59249 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:29:42 ; Search time 11371 Seconds
(without alignments)
2160.608 Million cell updates/sec

Title: US-09-316-048-5
Perfect score: 3506
Sequence: 1 acttcctgcggcgctgcggg.....taatactaaaaaaaaaaaaa 3506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp14:*
121: em_estp15:*
122: em_estp16:*
123: em_estp17:*
124: em_estp18:*
125: em_estp19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est67:*
136: gb_est68:*
137: gb_est69:*
138: gb_est70:*
139: gb_est71:*
140: gb_est72:*
141: gb_est73:*
142: gb_est74:*
143: gb_est75:*
144: gb_est76:*
145: gb_est77:*
146: gb_est78:*
147: gb_est79:*
148: gb_est80:*
149: gb_est81:*
150: gb_est82:*
151: gb_est83:*
152: gb_est84:*
153: gb_est85:*
154: gb_est86:*
155: gb_est87:*
156: gb_est88:*
157: gb_est89:*
158: gb_est90:*
159: gb_est91:*
160: gb_est92:*
161: gb_est93:*
162: gb_est94:*
163: gb_est95:*
164: gb_est96:*
165: gb_est97:*
166: gb_est98:*
167: gb_est99:*
168: gb_est100:*
169: gb_est101:*
170: gb_est102:*
171: gb_est103:*
172: gb_est104:*
173: gb_est105:*
174: gb_est106:*
175: gb_est107:*
176: gb_est108:*
177: gb_est109:*
178: gb_est110:*
179: gb_est111:*
180: gb_est112:*
181: gb_est113:*
182: gb_est114:*
183: gb_est115:*
184: gb_est116:*
185: gb_est117:*
186: gb_est118:*
187: gb_est119:*
188: gb_est120:*
189: gb_est121:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	ID		
1	1757.8	50.1	1801	146	U69197	U69197 U69197 Soar
2	701	20.0	1018	136	BE867390	BE867390 601442470
3	646.2	18.4	1036	137	BE889880	BE889880 601512132
4	621.2	17.7	724	97	AW952360	AW952360 EST364430
5	617.2	17.6	752	137	BE894295	BE894295 601437666
6	606.4	17.3	821	107	BE379687	BE379687 601159257
7	604	17.2	609	107	BE380123	BE380123 601159257
8	603.8	17.2	686	110	BE613972	BE613972 601504093
9	602.2	17.2	674	38	AV707406	AV707406 AV707406
10	594.4	17.0	616	111	BE674157	BE674157 7d76a04.x
11	588.2	16.8	594	40	AW162002	AW162002 av72c03.x
12	571.4	16.3	612	38	AV695897	AV695897 AV695897
13	565.4	16.1	749	110	BE617223	BE617223 601441906
14	555	15.8	609	136	BE810330	BE810330 MR4-PT005
15	551.6	15.7	615	134	BE082712	BE082712 RC2-BT064
16	550	15.6	550	107	BE348298	BE348298 hw16b01.x
17	548	15.6	564	40	AW163206	AW163206 au97f05.x
18	546.4	15.6	829	106	BE300155	BE300155 600944602
19	542.4	15.5	584	136	BE813737	BE813737 RC1-BN003
20	534	15.2	673	135	BE781223	BE781223 601469159
21	531.2	15.2	536	136	BE837505	BE837505 RC2-FN009
22	531	15.1	866	136	BE868328	BE868328 601443887
23	530.2	15.1	573	3	AA206573	AA206573 zq51f08.r
24	517.8	14.8	910	138	BE966164	BE966164 601660093
25	513	14.6	543	134	BE089861	BE089861 RC5-BT070
26	511.4	14.6	513	13	AA889669	AA889669 ak54a04.s
27	510.4	14.6	512	27	AI983007	AI983007 wt46g02.x
28	508.8	14.5	852	137	BE872137	BE872137 601446238
29	500.6	14.3	673	109	BE541462	BE541462 601067913
30	497.4	14.2	633	89	AW320695	AW320695 uc22b10.y
31	488.2	13.9	569	3	AA191622	AA191622 zp81g10.s
32	487	13.9	516	25	AI819766	AI819766 wj42a10.x
33	483.4	13.8	485	25	AI802592	AI802592 wfi6a03.x
34	478.4	13.6	480	134	BE042598	BE042598 ho26e03.x
35	475	13.5	494	89	AW351909	AW351909 RCO-CT020
36	470.4	13.4	509	142	N31181	N31181 yx64f02.r1
37	470	13.4	517	7	AA427366	AA427366 zw33c05.r
38	470	13.4	517	92	AW580332	AW580332 IL2-HT044
39	467.4	13.3	703	137	BE888420	BE888420 601514177
40	457	13.0	724	38	AV700494	AV700494 AV700494
41	455.2	13.0	561	141	H16250	H16250 ym13d10.s1
42	451.8	12.9	455	96	AW874257	AW874257 hp96g04.x
43	450.4	12.8	452	17	AI190464	AI190464 qd38b02.x
44	446.4	12.7	586	141	H11397	H11397 ym13h07.s1
45	445.8	12.7	524	96	AW877286	AW877286 MR4-PT005

ALIGNMENTS

RESULT 1
U69197 U69197 1801 bp mRNA EST 27-OCT-1999
LOCUS U69197 Soares infant brain 1N1B Homo sapiens cDNA clone 22368, mRNA
DEFINITION U69197 Soares infant brain 1N1B Homo sapiens cDNA clone 22368, mRNA
ACCESSION U69197
VERSION U69197.1 GI:2739420
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1801)
AUTHORS Volorio,S., Simon,G., Repetto,M., Cucciardì,M., Banfi,S., Borsani
,G., Ballabio,A. and Zollo,M.
TITLE Sequencing analysis of forty-eight human image cDNA clones similar
to drosophila mutant protein
JOURNAL DNA Seq. 9 (5-6), 307-315 (1998)
MEDLINE 99452388
COMMENT Contact: Zollo, Massimo
Telethon Institute of Genetics and Medicine
Via Olgettina 58, Milan, MI 20132, Italy
Email: zollo@tigem.it.
Location/Qualifiers
1.1801
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p13.1"
/clone="22368"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGGAGAATTCGGCGCCGACGAAATTTTCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bernaldo."

BASE COUNT 503 a 409 c 386 g 503 t
ORIGIN

Query Match 50.1%; Score 1757.8; DB 146; Length 1801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1735 gctccgaatcctgccaaaggccacgtaactgcatatagcccgagagtggtgatggg 1794
|||||
Db 5 GCTCCGAATCCTGCCAAGGCCACGGTAAC TGCCATGATAGCCCCGAGAGTGTGTATGGG 64

QY 1795 ggcacctgcgccacagccgagacattttaaagaataacatctctcagccacgtacc 1854
|||||
Db 65 GGCACCTCGCCACAGCCGAGACCATTTTAAAGAAATACATCTCTTCAGGCCACGTACCC 124

QY 1855 catgacctctcacgagacctctgagcaactgacatcttccagagtcceaggtatc 1914
|||||
Db 125 CATGGAACCTCTCAGAGACCTCTGAGCAACTGAGATATCTTCCAGAGTCCAGGGATTTC 184

QY 1915 caggttgaaatacaagacttccccaanaaacaacaagaacgaatttgtatctcttacaat 1974
|||||
Db 185 CAGGTTGAATACAAAGACTTCCCAAAAACAACAAGCAAGCAATTGTATCTCTTATCAAT 244

QY 1975 tgcctctctcagccacctctgatacggcatgtatcgcaaggatgtgagtcctgcat 2034
|||||
Db 245 TGCCTCTCTCAGCCACCTCTGATCAGCCATGTGATCGGCAAGGATGTGGAGTCTCTCCAT 304

QY 2035 gatatgctgcgctgaaacattaaagttgctgtctgagtgtgagcaacaagaatcacagag 2094
|||||
Db 305 GATATGGCTGCGCTGAACATCTTAAAGTTGCTGTGAGTTGGACCAACAAGATGACAGAG 364

QY 2095 atgccaagaacaggaacggaacaaatgtctgtgtgtggaaggtgtgaaacctttctgac 2154
|||||
Db 365 ATGCCAAGAACAGGAACGACCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGC 424

QY 2155 catgaaccattataaatcccaacatatatactgaaataactgaaactgcttggaaatt 2214
|||||
Db 425 CATGAACCATTTATAAATCCCAACATATATACTGAAATACTGAAACTGCTTGAAAAATT 484

QY 2215 tggaaattctgtatacctccagatgagccgagagacacggttggtaagagatgtgagcagca 2274
|||||
Db 485 TGGAAATTCTGATACCTCCAGTGGGCCGAGAGACACAGGTGGGTAAGATGTGGCAGCA 544

QY 2275 gcagggagaacaacagaacaagaaggcgctgtgccc-ggctgactgtgtcgggt 2333
|||||
Db 545 GCAGGGAAGACAACAGAAACACAGAGGCGGCTGTGGCGGGCTGGACTGTGTGGCGGT 604

QY 2334 ttgttgtatggccactcgtgtacttggcggtccctacgcaatagcagctgtctgtggg 2393
|||||
Db 605 TTGTTGTGATGGCCACTCGGTGACTGGCGGTCCCTACGCATAGCAGCTGCCGTGGGG 664

QY 2394 aagaaggctgcccagccagcagctgttctccggygacacccagcagatccacacctgggca 2453
|||||
Db 665 AAGAAAGGCTGCCACGACCAAGCTGTCTCCGGGACACACAGATCCACACCCCTGGGCA 724

QY 2454 cctccgtgttggctcttcttccctgtgtgaaagaagaaacgycacgacccctctc 2513
|||||
Db 725 CCTCCGTGTGTGCTCTTTTTCCTCTGTGTGAAGAAGAAACGGCACGACCCCTTCTC 784

QY 2514 aagctgctcaactcagacacatitgggacaacaccttgacagccatgcccagagagagcct 2573
|||||
Db 785 AAGCTGGCTCACTCAGACACATTTGGACAACACCTTGACAGCCATGCCAGAGAGAGCCT 844

QY 2574 ttgaccggccccagagctaaagaaccagagaaatcaatgcttccctactcagctgtac 2633
|||||
Db 845 TTGACCGGCCCCAGAGCTAAAGACACAGAAATCAATGCTTCCCTACTCAGCGTGAC 904

QY 2634 ccaacttctctagtgtgtgccagcgccccaccacctctcgtcagtaaccacaccatccact 2693
|||||
Db 905 CCAACTTTTCTAGTGTGCGCACGGGCCCAACCACTCCTGCGAGTACCCACACCATCACCACT 964

QY 2694 gcttctctccaacagatgactgtatctttagttcattatcttctttagttagatg 2753
|||||
Db 965 GCTTCTCTTCCACAGTGAATCTGTATTCTTAGTTTCATTATTCTTTTGATTGATATG 1024

QY 2754 acactataaaaaatttcatttggaaattctcaattgtatctagttlaaatagcacagt 2813
|||||
Db 1025 ACACATATATAAAATTTTCATTGTGAGATTCTCAATTGTATCTAGTTAAATAGCACAGTT 1084

QY 2814 tggaaactgtctgagactgacttatacctaatactaacgcgacaagaatcatatcatcgt 2873
|||||
Db 1085 TGGAAACTGTCTGAGACTGACTTATCAATATCTTAACCGACAAGATCATATCCATGT 1144

QY 2874 gtatgtgttagacatttltatttcaattgactaacccagagacgttccagtgatgcaat 2933
|||||
Db 1145 GTATGTGGTTAGACATTTTATTTCAATTGACTAACCAGAGACGTTTCAGTGATGCAAT 1204

QY 2934 tgtgtccctctgtgttcagctgaaacagtcctgacttccaacaaaccttgaaatagctc 2993
|||||
Db 1205 TGTGTGCCCTCTGTGTTCAAGCTGAACAGACTCCTGGACTTTCAAAAACCTTGAATAAGTCTC 1264

QY 2994 ccaagttgtataaaatttgacaattttaggaattttaaactttagatgatacatlgtgtcc 3053
|||||
Db 1265 CCACAGTTGTATAAATTGGACAATTGGAATTTAACCTTTAGATGATCATTTGGTTCC 1324

QY 3054 attttatcatttattatttlttgaatgcaaacagagactaaatgaaactttagatctc 3113
|||||
Db 1325 ATTTTATTTCAATTTTATTTTGTAAATGCAAAACAGAGACTTAAATGAACCTTGATCTCT 1384

QY 3114 gttttaaagattatlaaaaaaacatgtgtatctatacatataggtctctgagacttagct 3173
|||||
Db 1385 GTTTTAAAGATTATTTAAAAACATTTGTATCTATACATATGGCTCTTGAGGACTTAGCT 1444

QY 3174 ttcaactacactacaggaatatgatctcctcatgtagtcataataaacctgcagagtgatctc 3233
|||||
Db 1445 TTCACTACACTACAGGATATGATTCATGTAGTCCATATATAACCTGCAGAGTGATTTTC 1504

QY 3234 cagaagtgcctgataactgttlaatatcaatctccatttagggctgaaaaagaatgacctagctt 3293
|||||
Db 1505 CAGAGTGTGATACCTGTTAATTACATCTCCATTAGGGCTGAAAAAGAAATGACCTAGCTTT 1564

QY 3294 ctgtatacagctgtgtgtctttagtgtgttactgtacacagaagtggtgtgcaactga 3353

[illegible]

RESULT	2
LOCUS	BE867390
DEFINITION	BE867390 1018 bp mRNA EST 27-SEP-2000 601442470F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846689 5', mRNA sequence.
ACCESSION	BE867390
VERSION	BE867390.1 GI:10316166
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1018)
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	

Tel: (301)496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM546 row: 1 column: 18
 High quality sequence stop: 654.

FEATURES	SOURCE
Location/Qualifiers	1. 1018
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:3846689"	
/clone_11b="NIH_MGC_65"	
/tissue_type="adenocarcinoma"	
/lab_host="DH10B (phage-resistant)"	
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."	
	"
295 a	267 c 262 g 194 t
BASE COUNT	
ORIGIN	

Query Match	20.0%;	Score 701;	DB 136;	Length 1018;
Best Local Similarity	90.0%;	Pred. No. 2.6e-163;		
Matches 809; Conservative	0;	Mismatches 80;	Indels 10;	Gaps 5;

QY	1280	tggtccagatccagcagycaaaaaagagaaagagccagagtlacacgctcctcacagagc	1339
Db	1	TGGCCAGATCCAGCAGGCAAAAAAGAGAGAGAGCCAGAGTACACGCTCTCTCACAGAGC	60
QY	1340	gaagcctcccgcgccgcagggagtltygatgcaggtgaagtltygaaaccacacatcag	1399
Db	61	GAGGCTCCCGCGCGCCGAGGAGTTTGATGCAGGTGAAGGTTGGAACACACACTGCAG	120
QY	1400	aaggaacggygcaccaacaagaagtlgyccaagcgcaatgcagccgagaacatgctggaga	1459

Db	121	AAGGAACGGGACCAACAGAAGGTGGCCAAAGCGCAATGCAGCCGAGAACATGCTGGAGA	180
QY	1460	tccttggtttcaaaagtcctccgcagcgcagcccccacaaacccgcactcaagtlcagaggaga	1519
Db	181	TCCTTGCTTCAAAAGTCCCGCAGCGCCAGCCACCAAAACCCGCACTCAAGTCAGAGAGA	240
QY	1520	agacacccataaagaacaccaggggatgtgaagaanaagtaacctttttgaaacctgtctg	1579
Db	241	AGACACCCCATTAAGAAACCAAGGGGATGGAGAGAAAGTAACCTTTTGTGAACCTGGCTCTG	300
QY	1580	gggatgaaaatgggactaagtaataaaggatgagttcaggatgacctatcctaagtcac	1639
Db	301	GGGATGAAATGGGACTAGTAATAAGAGGATGAGTTCCAGGATGCTTATCTAAGTCATC	360
QY	1640	agcagctgcctgtgnaattctcccatgtgtcccgaggtgcgccagcgtgtaggagltta	1699
Db	361	AGCAGCTGCCCTGCTGGAATTCTTCCCATGCTGCCGAGGTGCCAGGCTGTAGAGGTTA	420
QY	1700	gtcaagagacatcacaccaaaagattttaccagggcagctccgaatcctctgccaagccaag	1759
Db	421	GTCAGGACATCACACCAAGAATTTTACCAGGGCAGCTCCGAATCCTGCCAAGGCCACGG	480
QY	1760	taactgccatgtatagcccgagagttgtgtatgaggggcaacctgcgccacagccgagacca	1819
Db	481	TAACTGCCATGATAGCCCGAGAGTTGTGCTATGGGGCAACCTCGCCACAGCCGAGACCA	540
QY	1820	ttttaagaataacatctcttcagggccacgylacccccatygacctctcacgagaccctctg	1879
Db	541	TTTTAAGAATAACATCTCTTCAGGCCACGTAACCCATGGACCTCTCACGAGACCCTCTG	600
QY	1880	agcaactggaactatcttccagaggtccaggaatccagglttgaatacaaaacttcccca	1939
Db	601	AGCAACTGGACTATCTTCCAGAGTCCAGGGGATTCAGGTGAATACAAAGACTTCCCCA	660
QY	1940	aaaacaacaagaacgaattgtatct--ctatcaattgtctcctcagccaactctga	1996
Db	661	AAATCACCACAAGAAGCAATTCGGTATCTCCTTATCAATGGCTCCCTCAGCCACTCTGAT	720
QY	1997	tcagccatggtatcggcaagatgtgagttcctgcca--tgatatgctgagctgaaca	2053
Db	721	CCAACATGATGTTTCGGCAGGTTGTGGGAATCCTGGCAATGATCATGGGTGGGCTTGACA	780
QY	2054	tcttaaaag-ttgtctgtcctgagltgacccaacaaa-gtacagagatgccaagaacagga	2111
Db	781	TCTTTAAGATGGCTGGCGGAAGTGGACCCACAAAGCTCAAAAAATTTCCACGAACGGGAA	840
QY	2112	cggaccaatgtctgtgtgtggg--aggtgctgaacctttctgcccattgaaccattata	2168
Db	841	GGGACCCAGTTCGGTGTGTTTGGCAGACAGGTCTCGAACCAATTTTGGGCTGTAACCCATTAA	899

RESULT	3			
BE889880				
LOCUS				
DEFINITION				
ACCESSION	BE889880	1036 bp	mRNA	EST
VERSION	601512132F1	NIH_MGC_71	Homo sapiens	CDNA clone IMAGE:3913590 5',
KEYWORDS	mRNA sequence.			
SOURCE	BE889880			
ORGANISM	BE889880.1	GI:10347645		
	EST.			
	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 1036)			
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D.			
	Tel: (301) 496-1550			
	Email: Robert_Strausberg@nih.gov			
	Tissue Procurement: ATCC			
	CDNA Library Preparation: Life Technologies, Inc.			

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>

Plate: LNCM720 row: p column: 07
High quality sequence stop: 706.

Location/Qualifiers
1..1036

FEATURES
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3913590"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 296 a 230 c 220 g 290 t
ORIGIN

Query Match 18.4%; Score 646.2; DB 137; Length 1036;
Best Local Similarity 91.5%; Pred. No. 1.1e-149;
Matches 740; Conservative 0; Mismatches 63; Indels 6; Gaps 5;

QY 2579 cggccccagagctaaagcaccagagaataaactgcttcctactcagcgtgacccaac 2638
Db 1 CGGCCCCAGAGCTAAAGCACACAGAGAAAATCAATGCTTCTACTCAGCGTGACCCAAC 60

QY 2639 ttctctagtgccagggccccaccctcctgcagtlaccacacacacactgctt 2698
Db 61 TTTCTAGTGTGCCAGGGCCCCACACACCTCTGCAGTACCCACACCATCACTGCTTT 120

QY 2699 ctcttccaacagtgatcgtatcttctagtttcatatttcttcttgatgatgacact 2758
Db 121 CTCTTCCAACAGTGATCTGTATCTTAGTTTCATTTTCTTTTGATGATATGACACT 180

QY 2759 atataaaatttcaattgagaatttctcaattgtatctagtttaaatagcacagtttgaa 2818
Db 181 ATATAAATTTTCATTTGAGAAATTTCTCAATGTATCTAGTTAAATAGCACAGTTTGAA 240

QY 2819 actgtctgagactgacttatacaataatctaacgcacaagaatcatatcatgtgtatg 2878
Db 241 ACTGTCTGAGACTGACTTATCAATAATCTAACCGACAAGATCATATCATGTGTATG 300

QY 2879 tggttagacattttatttatttactgactaacccagagacgttccagtgtgcaaatgtgt 2938
Db 301 TGGTTAGACATTTTATTTCAATTGACTAACCCAGGACAGTTTCAGTATGCAATGTGT 360

QY 2939 gccctctgtgtcagctgaacacagctcctgagacttccaacacccctgaataagctccaca 2998
Db 361 GCCCTCTGTTGAGCTGAACAGTCTCTGAGCTTTCAAAAACCTTGAATTAAGTCTCCACA 420

QY 2999 gtgtatataaattgacaatttaggaattttaaactttagatgatcatgtgttccattt 3058
Db 421 GTTGTATTAATTTGACAAATTTAGGAATTTTAACCTTTAGATGATCATTTGTTCCATTT 480

QY 3059 taattcatttatttctgttaatgcaaacagagacttaaatgaactttgactctgtttt 3118
Db 481 TAATTTCATTTTATTTTGTAAATGCAAAACAGAGACTTAAATGAACCTTGAATCTCTGTTT 540

QY 3119 aaagatatataaaacattgttatctatacatatgctcttgagacttagcttccac 3178
Db 541 AAAGATTATTAAAAACATTTGTATCTATACATATGGCTCTTGAGAGACTTAGCTTTCA - 599

QY 3179 taacactacagatatgatctcc-atgtagtcataataaacctgcagagt--atttcca 3235
Db 600 TACACTACAGGATATGATCTCCCAATGTAGTCCATATAAACCTGCAGAGTGGATTTTCCAG 659

QY 3236 gagtgcctgatact-gttaattacatctccatagggctgaaaagaatgacctacgttcc 3294
Db 660 AAGTGTCTGATACTGTTCAATTTACTCTCCATTAAGGCTGAAGAAAGATGACCTAGTTCTG 719

QY 3295 tgratacagctgtgtgtctttgatgtgtgttactgtacacagaagtgtg-tgcactga 3353
Db 720 TATTAAGTGTGTGTGTGTGACAGGTTCGGGTTACTGTGTGAACAGACAGAAGTGTTCCTCGA 779

QY 3354 ggcctgcgtgtgtgtccglatggaacc 3382
Db 780 GGGTCTGGATGTGGGCCGCTTGATATAC 808

RESULT 4
AW952360
LOCUS
DEFINITION AW952360 724 bp mRNA EST 01-JUN-2000
ACCESSION AW952360
VERSION AW952360.1 GI:8142042
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS
1. (bases 1 to 724)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspari,R., Gay,C., Holt
, I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

TITLE
JOURNAL
COMMENT
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 38
Seq primer: Reverse.

FEATURES
source
Location/Qualifiers
1..724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGB"
/note="Vector: pBluescriptSkm"

BASE COUNT 198 a 179 c 189 g 158 t
ORIGIN

Query Match 17.7%; Score 621.2; DB 97; Length 724;
Best Local Similarity 95.1%; Pred. No. 1.7e-143;
Matches 685; Conservative 0; Mismatches 28; Indels 7; Gaps 4;

QY 1739 cgaatcctgccaagcgcaagcgttaactgtccatgatagcccgagagttgtgtatggggca 1798
Db 1 CGAATCCTGCCAAGGCCACGGTAACTGCCATGATAGCCGAGAGTGTGTATGGGGCA 60

QY 1799 cctgcgccacagccgagacattttaaagaataacatctcttcaggccacgtaccccatg 1858
Db 61 CCTGCCACAGCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCACGTACCCCATG 120

QY 1859 gacctctacagagaccctctgagcaactgagactatcttcagaggtccagagatccagg 1918
Db 121 GACCTCTCACGAGACCCCTCTGAGCAACTGAGACTATCTTCCAGAGTCCAGGATTCAGG 180

QY 1919 ttgaataaaagacttccccaaaaacaagaagaattgtatctctatacaattgct 1978
Db 181 TTGAATACAAGAGACTTCCCAAAAACAACAAGAAGCAATTTGTATCTTATCAATTTGCT 240

QY 1979 cctctcagccacactctgatcagccatggtatcggaagagtgtgagtcctgcatagata 2038
Db 241 CCTCTCACGCCACCTCTGATCAGCCATGTFATCGGCAAGAGATGTGAGATCCTGCCATGATA 300

QY 2039 tggctgcgtgaacatcttaagtgtcgtcgtgagtgaggaaccaacaagaagtacagatgc 2098
Db 301 TGGCTGCCCTGAACATCTTAAGTTGCTGTCGTGAGTTGACCAACAAGATACAGAGATGC 360

Query Match	Best Local Similarity	Score	DB	Length	752;
Matches 684; Conservative	94.2%;	Pred. No. 1.6e-142;	Mismatches 38; Indels 4; Gaps 4;		

OY	1247	aataatgcccagggtatcaatccgaattagccgacttggccccagatccaggcagaacaaagg	1306
Db	1	AAATATGGCCAGGGGATCAATCGATTGACCGACTGGCCCCAGATCCAGCAGGCCAAAAAGG	60
OY	1307	agaaggagccagagtacaagcttcctcacagagcgaggcctccgcgccgagggagtttg	1366
Db	61	AGAAGGAGCCAGAGTACACAGCTTCCTCACAGAGCGAGGCCTCCCGCCGACGAGGTTCG	120
OY	1367	tgatgcaggtgaaggttggaaccacacctgcagaaggaacggygcaccaacaagaagtgc	1426
Db	121	TGATGCAGGTGAAGGTTGGAAACCACACTGTCAGAGAAGAACCGGCACCAACAAGAAGTG	180
OY	1427	ccaagcgcgaatgcagccgagaacatgctygagatccttgtttcaaagtcccgacgyc	1486
Db	181	CCAAGCGCAATGCAGCCGAGACATGCTGGAGATCCTTGTTCAAAGTCCCGCAGGCGC	240
OY	1487	agcccaaccaaacccgcactcaagtcatagagagagaagacaccataaagaacagggatg	1546
Db	241	AGCCCAACCAACCCGCACTCAAGTCAGAGGAGAAGACACCAATAAAGAAACCA-GGCAATG	299
OY	1547	gaagaaaagtaaccctttttgaaacctgctctggygatatgaaaaatyggactagtataaag	1606
Db	300	GAAGAAAAGTAACCTTTTGTGAACCTGGCTCTGGGGATGAAATGGGACTAGTAATAAG	359
OY	1607	aggatgagtcagagatgccttatctaagtcatacagcagctgctctgtygaattcttcca	1666
Db	360	AGCATGAGTTCAGGATGCTTATCTAAGTCATCAGCAGCTGCTGGAATTCCTTCCA	419
OY	1667	tggtgccgagatgcgccagctgtagaggttagtcaagagacalcacaccaaatlta	1726
Db	420	TGGTGCCCGAGGTGCGCCAGGCTGTAGAGGTTAGTCAAGGACATCACACCAAGATTTTA	479
OY	1727	ccaagggcagctccgaatcctctgccaagggcaacgtaactgcatatagcccgaggtgc	1786
Db	480	CCAGGACAGCTCCGATCTCTGCCAAGGCCACGGTAACCTGCATGATAGCCGAGAGTTGT	539
OY	1787	tgtatggggcaccctcgcccacagccgagaccattttaa-gaataacatctctcagc	1845
Db	540	AGTATGGGGGCACCTCGCCCCACAGCCAGATCATTTTAAACGAATAACATCTCTTCAGGC	599
OY	1846	cacgtaccccatltagacctctcaagagaccctctgagcaactgactatcttccagagtc	1905
Db	600	CACGTACCCCATGGA-CTCTCAGAGATCCTCTGAGCAA-TGCACTATCTTCCAGAGTC	657
OY	1906	caggagatccaggttgtaatacaagaacttccccaaaaacaagaagaacgaattgtatct	1965
Db	658	CAGGATTCGGGTTGATACAAAGACTTTCCCAAAAAAAAAACAGAGAATTGTTCTCTATCA	717
OY	1966	cttacc 1971	
Db	718	ATGCTC 723	
RESULT	6		
LOCUS	BE379687/c		
DEFINITION	601159257T1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510806 3'		
ACCESSION	BE379687		
VERSION	BE379687.1 GI:9325052		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 821)		
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert_Strausberg@nih.gov		


```

QY 2673 agtaccacacatcaaccactgtcttctcttccaaagtgatctgtattcttaagttcat 2732
    |||
Db 121 AGTACCACACACATCACCACTGCTTCTCTCCACAGTGATGTATTCTTAGTTTCAT 180
QY 2733 tattctcttgattgatataagacataataaatttcaattgagaatttctcaattgt 2792
    |||
Db 181 TATTTCTTTGATTGATATGACACTATATAAATTTTCATTGAGAAATTTCTCAATTGT 240
QY 2793 atctagtaataagcacagtttggaactgtctgagactgacttatacaataatcaac 2852
    |||
Db 241 ATCTAGTTAAATGACACAGTTTGAAACTGTCTGAGACTGACTTTATCAATAATCTAAC 300
QY 2853 cgacaaagatcatatccatgtgtatgtgttagacattttattcattgactgaaccag 2912
    |||
Db 301 CGACAAAGATCATATCCATGTGTATGTGTTAGACATTTTATTTCAATTGACTAACCCAG 360
QY 2913 gacagtttcagtgatgcaaatgtgtgcccctgtgttcagctgaaacagtcctgactt 2972
    |||
Db 361 GACAGTTTCAGTGAATGCAATGTGTGCCCCCTGTGTTACAGCTGAACAGTCCCTGACTTT 420
QY 2973 caaaaacctgaataagctctcccacagttgtataaattggaacatttaggaattttaa 3032
    |||
Db 421 CAAAAACCTTGAAATAGTCTCCACAGTGTATATAATTGGACAAATTTAGGAATTTTAAAC 480
QY 3033 tttagatgatcattgtgttccatttatttatttatttatttatttatttatttattt 3092
    |||
Db 481 TTTAGATGATCATTTGTTCCATTTTATTTTCAATTTTATTTTGTAAATGCAAAACAGGA 540
QY 3093 cttaaatgaacttgatcctctgttttaagaattataaataaacaattgtgtatctata 3152
    |||
Db 541 CTTAAATGAACCTTGATCTCTGTTTAAAGATTATTAAAAACATTTGTATCTATACAT 600
QY 3153 atgg 3156
    |||
Db 601 ATGG 604

RESULT 8
BE613972 686 bp mRNA EST 24-AUG-2000
LOCUS 601504093F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5',
DEFINITION mRNA sequence.
ACCESSION BE613972 GI:9895569
VERSION BE613972
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 686)
REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM700 row: j column: 24
High quality sequence stop: 675.
FEATURES
source location/Qualifiers
1..686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3905783"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"

```

```

/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 226 a 156 c 196 g 108 t
ORIGIN
Query Match 17.2%; Score 603.8; DB 110; Length 686;
Best local Similarity 97.1%; Pred. No. 3.4e-139;
Matches 669; Conservative 0; Mismatches 12; Indels 8; Gaps 5;

QY 854 aacacgatgctctgccaagcggttgagatcctgcagaatgaagccctgcagaagagc 913
    |||
Db 1 AACACGATGCTGCTGCCAAAGCGTTGAGGATCTGCAGAAATGACCCCTGCCAGAGAGGC 60
QY 914 tgyaagtgaaatggaagaatccgaagaagaaatctcaataaatctgaataagtaag 973
    |||
Db 61 TGGAGGTGAATGGAAGAAGATCCGAAGAAGAAATCTCAATTAATCTGAATAATGTCAG 120
QY 974 tgttagatctgcaacttaaacggaactgtgctgtgaatttcagagtgccgagagagtg 1033
    |||
Db 121 TGTTTGAGATGCACTTAACGGAACCTGCTGTGAATTTTGAGGTGCGCCGGAGAGTG 180
QY 1034 gccacccccacatgaagaacttgtgacaaaggttgcgttgaggaggttggtggggaag 1093
    |||
Db 181 GCCCACCACACATGAAGAAC-TTGTGACCAAGGTTTCGGT--GGGAGTTGTGGGGGAG 237
QY 1094 gtgaagggaagaagcaagaagatttcaagaanaaatgcccgcatagtctgttcttgaggagc 1153
    |||
Db 238 GTGAAGGGAAGAAGCAAGAAGATTTCAGAAGAAAATGCCGCCATAGCTGTCTTGAGAGAGC 297
QY 1154 tgaagaagtlaccgccccctgctgctgtgaacgagtaagtagtaaatcaaaaaaagaaa 1213
    |||
Db 298 TGAAGAAGTTACGCCCTGCTGCTGACGTTGAACGAGTAAAGCTTAGAATCAAAAAGAAA 357
QY 1214 caaaaccatagtcaagccacagacaaagcccaagaatatgcccaggggatcaatccgatta 1273
    |||
Db 358 CAAAACCATAGTCAAGCCACAGACAAAGCCCAAGATATGGCCA-GGGATCAATCCGATT 416
QY 1274 gccgactgcccagatcagcagcgaagaaagagaagagcagagtagtacagctctca 1333
    |||
Db 417 GCCGACTGCCCCAGATCCAGCAGGCAAAAAGAGAAGAGAGCAGATACACGCTCCATCA 476
QY 1334 cagagcgagcctcccgccgcccagggaggttgatgacgtgaaggttgaaaccaca 1393
    |||
Db 477 CAGAGCGAGGCTCCCGCGCCGACGAGGAGTTGTGATGACAGTGAAGGTTGGAACACACA 536
QY 1394 ctgcagaagaagcgggcacacaaagaaggttgcccaagcgcaatgcaagccgagaacatgc 1453
    |||
Db 537 CTGCAGAAAGGAACGGGCACCAACAAGAGAGTGCCCAAGCCGCAAGCAGAGACATGC 596
QY 1454 tggagatccttgg-ttcaaaagtcgccgagcgagccaccacaaccgcactcaagtca 1512
    |||
Db 597 TGGAGATCCTTGTTTTCAAAAGTCCCGCAGGCGCAGCCACCAAA--CCGACTCAAGTCA 653
QY 1513 gaggagaagacacccatlaaagaaccagg 1541
    |||
Db 654 GAGGAGAAAGACCCCATTAAGAAACAGGG 682

RESULT 9
AV707406 674 bp mRNA EST 09-OCT-2000
LOCUS AV707406 ADB Homo sapiens cDNA clone ADBCMG03 5', mRNA sequence.
DEFINITION AV707406
ACCESSION AV707406
VERSION AV707406.1 GI:10724671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 674)
REFERENCE

```


AUTHORS Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA ADB clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1.674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADBCMG03"
/clone_lib="ADB"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 197 a 148 c 113 g 215 t 1 others
ORIGIN

Query Match 17.2%; Score 602.2; DB 38; Length 674;
Best Local Similarity 98.4%; Pred. No. 8.6e-139;
Matches 618; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 2510 tctcaagctggtcactcagacacattgggacaacacctgacagcagcatgcccagagag 2569
Db 1 TCTCAAGCTGGCTCACTCAGACACATTGGGACAAACCTGGACAGCCATGCCAGAGAG 60

QY 2570 gcccttgaccggcccccagagctaaagcaccagagaaatcaaatgtcttctactcagcg 2629
Db 61 GCCTTGGACCGGCCCCAGAGCTAAAGCACAGAGAAATCAATGCTTCTACTCAGCG 120

QY 2630 tgaaccacttttctagtgtgccacggcccccaccacctctctgagtlacccaccatcac 2689
Db 121 TGACCCCACTTTCTAGTGTGCCACGGCCCCACCACCTCTCTGCACTACCCACATCAC 180

QY 2690 caactgttctctccaacagatctgtatcttctagttcattatlttctttgattga 2749
Db 181 CACTGCTTCTCTCCAACAGTGAATCTGTATCTTAGTTCAATTATTTCTTTGATTGA 240

QY 2750 tatgacacataaaatttctattgagaatttctcaattgtatctagttaataagcac 2809
Db 241 TATGACACTATATAAATTTCAATTGAGAATTTCTCAATTGTATCTAGTTAAATGACAC 300

QY 2810 agttggaactgtctgagactgactttatcaataatctaacccgacaagaatcatatcc 2869
Db 301 AGTTTGGAAACTGTCTGAGACTGACTTTATCATATATTAACCGACAAGATCATATCC 360

QY 2870 atgtgtatgtgttagacatttatttctattgactaacccagagacagtttcagtatgc 2929
Db 361 ATGTGTATGTGGTTAGACATTTTATTTCATTGACTAACCCAGGACAGATTTCAGTATGC 420

QY 2930 aaatgtgtccctctgtgtcagctgaaacagtcctggaacttcaaaaaacctgataag 2989
Db 421 AAATTTGTGCTCTGTGTTCACTGAAACAGTCTCTGACTTTTCAAAAAACCTTGAAATAG 480

QY 2990 tctccacagttgtataaattggacaatttaggaattttaaactttagatgatttg 3049
Db 481 TCTCCACAGTTGTATAAATTGGCAATTTAGGAATTTTAACTCTAGATGATCATTTGG 540

QY 3050 ttccatttatttcaatttatttattgttaatgcaaacagacttaaatgaacttgat 3109
Db 541 NTCCA-TTTTATTTTCATTCTTATTTTGGTAATGCAAAACAGCACTTAAATGAACTTTGAT 599

QY 3110 ctctgtttaaagattattataaaacat 3137
Db 600 CTCGTGTTTAAAGAGTATTAACACGAT 627

RESULT 10
BE674157
LOCUS BE674157 616 bp mRNA EST 08-SEP-2000
DEFINITION 7d76a04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3278862 3' similar to TR:O95793 O95793 STAUEN PROTEIN.; mRNA sequence.
ACCESSION BE674157
VERSION BE674157
KEYWORDS BE674157.1 GI:10034698 EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 616)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 469.

FEATURES
source
1.616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3278862"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 179 a 153 c 139 g 141 t 4 others
ORIGIN

Query Match 17.0%; Score 594.4; DB 111; Length 616;
Best Local Similarity 98.4%; Pred. No. 7.3e-137;
Matches 598; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1519 aagacacccaataaagaaccagggagtgaagaagaataacctttttgaacctgctct 1578
Db 8 AAGACACCCATTAAGAAACACAGGGAGTGAAGAAAGTAACCTTTTGAACCTGGCTCT 67

QY 1579 gggagtgaanaatggygactagtaataaagagatgagttcagagatgccttatcgaatcat 1638
Db 68 GGGGATGAATAATGGGACTAGTAAATAAGAGATGAGTTCAAGATGCCCTTATCTAAGTCAT 127

QY 1639 cagcagctgctctgtgaattcttcccatgtgcccagagtcgcccagctgttagaggtt 1698
Db 128 CAGCAGCTGCTGCTGGAATTCCTCCATGCTGCCCGAGGTCGCCAGGCTGTAGAGATT 187

QY 1699 agtcaagacatcacaccaagaatttaaccaaggcagctccgaaatcctgccaagccacg 1758

REFERENCE 1 (bases 1 to 612)
AUTHORS Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,
Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GK- clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1. .612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCEF08"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 149 a 164 c 125 g 174 t
ORIGIN

Query Match 16.3%; Score 571.4; DB 38; Length 612;
Best Local Similarity 98.4%; Pred. No. 3.8e-131;
Matches 611; Conservative 0; Mismatches 1; Indels 9; Gaps 3;

QY 2352 ggtgacctggtccctcctacgaatagcagctgtctgtgggaagaaggctgccagcc 2411
|||||
Db 1 GGTGACCTGGCGGTCCCTACGCAATAGCAGCTGCTGTGGGAGAAGGGCTGCCAGCC 60

QY 2412 agctgttctcccggaacccagcagatccacacccctgtggcactccgtgtttgtctt 2471
|||||
Db 61 AGCTGTTCTCCCGGACACGACAGATCCACACCCCTGGGCACCTCCGTGTTGGTTT 120

QY 2472 ttttccctgtgtgaagaagaacggaagaccccttctaagctgtctcaactcagac 2531
|||||
Db 121 TTTTCCCTGTGTGAAGAAGAAGCGGACGACCCCTTCTCAAGCTGGCTCAGCTCAGAC 180

QY 2532 acatgggacaacccctgagcagccatgccagagagagcccttgaccggccccaagct 2591
|||||
Db 181 ACATGGGACAACCCCTGGACAGCCATGCCAGAGAGAGCCCTTGACCGGCCCCAGAGCT 240

QY 2592 aaaaagcaccagagaataatcaatgtctcctactcagcgtgacccaactttctagtgtc 2651
|||||
Db 241 AAAAGCACACGAGAAAATCAATGCTTCTACTCACTGACCGTGAACCACTTTCTAGTGTGC 300

QY 2652 caagggcccaaccacctctgtcagtaaccacaccatcaaccactgtcttcttccaacagt 2711
|||||
Db 301 CACGGCCCCACACCTCTCGAGTACCCACACCATCACCACTGCTTCTCTTCCAACAGT 360

QY 2712 gactgtattcttagtttcaattatttcttcttgatgatatagacaataataaatttc 2771
|||||
Db 361 GACTGTATTCTTAGTTTCAATTATTTCTTTTGAATTGATATGACACATATAT---TTTTC 416

QY 2772 atttgagaatttctcaattgtatctagttaaatagcacaglttgaaactgtctgagac 2831
|||||
Db 417 ATTGAGAAATTTCTCAATGTATCTAGTTAAATAGCACAGTTTGAAACTGTCTGAGAC 476

QY 2832 tgaacttatacaataatcaaccgacaagaatcatatccatgtgtatgtgttagacattt 2891
|||||
Db 477 TGACTTTATCAATAATCTAACCGACAAGATCATATCCATGTGTATGTGGTTAGACATTT 536

QY 2892 ttattcatgtactaacccaagacagtttcagtgatgcaaatgtgtgccctctgtgtca 2951
|||||

Db 537 TTATTTCATTGACTAACCACGACAGTTCAGTGATGC--ATTGTGTGCCCTCTGTGTCA 594

QY 2952 gctgaacagctcctgactt 2972
|||||
Db 595 GCTG--CAGTCTGTGACTTT 612

RESULT 13
BE617223
LOCUS BE617223 749 bp mRNA EST 24-AUG-2000
DEFINITION 601441906F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846265 5',
mRNA sequence.
ACCESSION BE617223 GI:9888161
VERSION BE617223.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 749)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM545 row: k column: 02
High quality sequence stop: 705.

FEATURES
source
1. .749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846265"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 244 a 166 c 223 g 116 t
ORIGIN

Query Match 16.1%; Score 565.4; DB 110; Length 749;
Best Local Similarity 95.1%; Pred. No. 1.2e-129;
Matches 714; Conservative 0; Mismatches 21; Indels 16; Gaps 12;

QY 835 aagacaagacagctgtcgaacaacgatgtctgtgccaaa-gcgttgagatcctgcagaa 893
|||||
Db 1 AAGACAAGACAGGCTGCGAACAACGATGCTGTGCCAAAGCGTTGAGGATCTGCAGAA 60

QY 894 tgaagccctgcagagagctggaagtgtaatggaagaaatccgaagaagaatactcaa 953
|||||
Db 61 TGAGCCCTGCCAGAGAGGCTGGAAGGTGAATGGAAGAAATCCGAAGAAGAAATCTCAA 120

QY 954 taatctgaataagtcaagtgatt-gagattgcactaaacggaacttgctgtgaatt 1012
|||||
Db 121 TAAATCTGAATAAGTCAAGTGTTTGAGATGTGCACTTAAACGGAACCTTGCCGTGAATT 180

QY 1013 tcgagtggtcccgagagagtgcccaaccacatgaagaacttgtgaccaagttcgg 1072
|||||
Db 181 TCGAGGTGGCCCCGGAGAGAGTGCGCCACCCACATGAGAAG-TTGTGACCAAGGTTTCGG 239

QY 1073 ttgggagtttgtggggaagtgaaaggaaagcaagaagatttcaagaanaaatgcg 1132
|||||

```

Db 240 TTGGGGAG-TTGTGGGGGAAGGTGAAGGGAAAAAGCAAGAGATTTCAAAAGAAATGCCG 298
QY 1133 ccatagctgtcttgaggagctgaagaagtaccgccccctgctgcagtgtgaacgagtaa 1192
Db 299 CCATAGCTGTCTTGAGGAGACTGAAGAAGTTAACCG-CCCTGCTCTCAGTTGAACGAGTTAA 357
QY 1193 agcctagaatcaaaaaaagaacaaacaccatagtcataagccacagacaagcccagaatatg 1252
Db 358 AGCCTAGATCAAAAAGAAACAAACCCATAGTCAAGCCACAGACAAGCCAGATATG 417
QY 1253 gccaggggatcaatccgatttagccgactgcccagatccagcagycaaaaaagagaag 1312
Db 418 GCCA-GGGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGGCCAAAAAGGAGAAG 476
QY 1313 agccagagtacacgctcctcacagaagcggcctcccgccgcagggaggttgatgc 1372
Db 477 AGCCAGAGTACACGCTCTCTCACAGAGCGAGG-CTCCCGCGCCGACAGGGAG-TTGTGATGC 534
QY 1373 aggtgaaggttgaaaccacactgcagaagaacgycaccaa-c-aagaagtgccaa 1431
Db 535 AGGTGAAGGTGGAAACCACTGCAGAGAACGGGCACCAACAAGAAGTGGCCAA 594
QY 1432 cgcaatgcagccgagacatgtgtgagatccttggtt--caaagtcgccgagcgagc 1489
Db 595 CGCAATGCAGCCGAGACATGCTGGAGATCCTTGTTCAAAAGTCCCGACAGGGCAGC 654
QY 1490 ccaccaaacccgactcaagtcaagagagaacacccataaagaaccagggatgtgaa 1549
Db 655 CAACAAACCGGACTC--AGTCAGAGAGAGACACCCATAA--GAACCGGGGATGGAA 709
QY 1550 gaaaagtaacctttttgaacctggtctg 1580
Db 710 GACACGTAACTTTTGGAACTGGCGCGGG 740
```

```

RESULT 14
BE810330 609 bp mRNA EST 21-SEP-2000
LOCUS BE810330 MR4-PT0052-220500-206-f06 PT0052 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE810330
ACCESSION BE810330.1 GI:10242524
VERSION BE810330.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 609)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
```

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=MR4-PT0052-220500-206-f06&f3=2000-05-22&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 589.

FEATURES
Location/Qualifiers

```

source 1. .609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="PT0052"
/dev_stage="Adult"
/note="Organ: pnet; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 173 a 153 c 145 g 138 t
ORIGIN
```

Query Match 15.8%; Score 555; DB 136; Length 609;
Best Local Similarity 97.9%; Pred. No. 4.5e-127;
Matches 594; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

```

QY 1586 aaatgagactgaataaagagatgattcagatgcttatctaatcatcagcagc 1645
Db 6 AACATGGGACTAGTAAT-AAGAGATGAG-TCAGGATGCCCTTATCTAAGTCATCAGCAGC 63
QY 1646 tgcctgtggaattctcccatgtgtgccgaggtgcgccagcctgtagagtagtcaag 1705
Db 64 TGCCGTCT-GAATTTCTTCCCATGTGTGCCGAGGTGCGCCAGGCTGTAGAGTTAGTCAAG 122
QY 1706 gacatcacaccaagaattttaccagggcagctccgaatcctgcgaagggccagcgttaactg 1765
Db 123 GACATCACACCAAGAATTTTACCAGGGCAGCTCCGAATCCTGCCAGGGCCACGGTAACTG 182
QY 1766 ccatgtagcccgagagtggtgtatgagggcacctcgccacagccgagaccatttaa 1825
Db 183 CCATGATAGCCCGAGAGTAGTGTATGGGGGCACCTCGCCACAGCCGAGACCATTTTAA 242
QY 1826 agaataacatctctcagggccagctaccatgacacctcacagaaaccttgagcaac 1885
Db 243 AGAATAACATCTCTTCAGGCCACGATGACCTCTCACAGAGACCCCTGTGAGCAAC 302
QY 1886 tggactatcttccagagtcagggatccaggttgtaatacaagaacttcccaaaaca 1945
Db 303 TGGACTATCTTTCACAGAGTCCAGGATTCAGGTTGAATACAAAGACTTCCCAAAAAACA 362
QY 1946 acaagaacgaattgtatctcttatcaattgtcctctcagccaacctctgacgcatg 2005
Db 363 ACAAGAACGAATTTGTATCTCTTATCAATGGCTCCTCAGCCACCTCTGATCAGCCATG 422
QY 2006 gatacgcaagatgtgagtcctcgtgccatgatattgctgcgtcgaacatcttaagtgc 2065
Db 423 GTATCGGCAAGAGATGTGAGTCTCTGCCATGATATGGCTGCGCTGAGACATCTTAAAGTTGC 482
QY 2066 tgtctgagttgaccaacaagaagtagagatgccaagaacaggaacggaacaaatgtctg 2125
Db 483 TGTCTGAGTAGGACCAACAAGTACAGAGATGCCAAGAACAGGAACGCAATGTCTG 542
QY 2126 tgtgtgaggtgtgaaacctttctgagcatgaaacattataaaatcccaatatata 2185
Db 543 TGTGTGGAGGTGTGAACCTTTTCTGGGCAATGAACCATTTATAAATCTCAACATATATA 602
QY 2186 ctgaaaa 2192
Db 603 CTGGCAA 609
```

```

RESULT 15
BE082712/c 615 bp mRNA EST 12-JUN-2000
LOCUS BE082712
DEFINITION RC2-BT0642-150200-012-a08 BT0642 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE082712
VERSION BE082712.1 GI:8473017
KEYWORDS EST.
SOURCE human.
```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE 1 (bases 1 to 615)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=tl2-RC2-BT0642-150
200-012-a08&tl3=2000-02-15&tl4=1)
Seq primer: puc 18 forward
High quality sequence start: 75
High quality sequence stop: 609.
Location/Qualifiers
1. 615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0642"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 110 a 159 c 153 g 193 t
ORIGIN
Query Match 15.7%; Score 551.6; DB 134; Length 615;
Best Local Similarity 97.4%; Pred. No. 3.2e-126;
Matches 571; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
QY 1100 ggaaaaagcaagaagattcaagaataatgccgcacatagctgttcttgagagctgaaga 1159
DB 615 GGAAGCAAGAGATTTCAGAGAAAAATGCCGCATAGCTGTTCTTGAGGAGCTGAAGA 556
QY 1160 agttaccgccccctgcctgcagtgaacgagtaaaagcctagaatcaaaaagaacaaac 1219
DB 555 AGTTACCGCCCCCTGCTGAGTAGAAGAGTAAGCCTAGAAATCAAAAAAGAAACCAAAC 496
QY 1220 ccatagtcaagccacagacagccagatatgycaggggatcaatccgattagccgac 1279
DB 495 CCATAGTCAAGCCACAGACAGACCAAGATATGGCCAGGGGATCAATCCGATTAGCCGAC 436
QY 1280 tggcccaagatccagcagcacaataaaggagaagggccagagtacacgctcctcacagagc 1339
DB 435 TGGCCCAAGATCCAGCAGCAGCAAAAAAGAGAAGAGCCAGAGTAGACGCTCTCACAGAGC 376
QY 1340 gaggcctcccgccgcccgaagggattgtgatgcaggtgaaggttggaaaccacactgcag 1399
DB 375 GAGGCTTCCCGCGCGCCGAGGAGTTGTGATGACAGGTGAAGGTTGAAACCACTGCAG 316
QY 1400 aaggaacggygcaaccaagaaggttggccaagcgcaatgcagccgaagaacatgtgaga 1459
DB 315 AAGGAACGGGCAACCAAGAAGGTGGCCAAGCGCAATGCAGCCGAGAACATGCTGGAGA 256

QY 1460 tccttgtttcaaagtcccccagcgcgcagcccccacaaacccgcactcaagtcagaggaga 1519
DB 255 TCCTTGTTTCAAGTCCCGCAGGCGCAGCCACCACCAACCCGCACCTCAAGTCAGAGAGA 196
QY 1520 agacaccataaagaagaaccagggatgtgaaagaataacctttttgaacctggtcttg 1579
DB 195 AGACACCCCAATAAGAAACCAGGGGATGGAAGAAAGTAACCTTTTGTGAACCTGGCTCTG 136
QY 1580 gggatgaataatgggactagtaataaagagatgagttcagatgcctatcctaagtcac 1639
DB 135 GGGATGAATAATGGGACTAGTAATAAGAGATGAGTTCAGGATGCCCTTATCTAAGTCATC 76
QY 1640 agcagctgcctgctggaattcttc-ccatggtgccgaggtcgccc 1684
DB 75 ACCATCTTCCTGCTGACTTCTTCACCATGTGTCGAGTTTCATCC 30

Search completed: April 5, 2001, 00:30:10
Job time: 53106 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:30:05 ; Search time 701.87 Seconds
(without alignments)
1876.520 Million cell updates/sec

Title: US-09-316-048-5
Perfect score: 3506
Sequence: 1 acttcctgcgcggcgtcgcg...taatactaaaaaaaaa 3506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36.*
1: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT.*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3502.8	99.9	3506	21	Z933322	Human stauflen cDNA
2	2914.8	83.1	3217	21	Z933323	Human stauflen cDNA
3	2896	82.6	3142	21	Z933324	Human stauflen cDNA
4	2879	82.1	3260	21	Z933327	Human stauflen cDNA
5	2866	81.7	3190	20	X90786	DNA encoding human
6	1256	35.8	2859	21	Z933328	Mouse stauflen cDNA
7	1107.2	31.6	1127	20	X39991	Prostate cancer as
8	798.8	22.8	946	20	X40002	Prostate cancer as
9	755.6	21.6	844	20	X40000	Prostate cancer as
10	753.2	21.5	773	20	X39992	Prostate cancer as
c 11	551	15.7	825	20	X40001	Prostate cancer as
c 12	501.4	14.3	794	20	X39993	Prostate cancer as

c	13	433.8	12.4	660	21	Z79981	Human colon cancer
c	14	412.2	11.8	727	21	Z80269	Human colon cancer
c	15	369	10.5	769	20	Z17411	Human gene express
c	16	327.8	9.3	708	21	Z80703	Human colon cancer
c	17	286.6	8.2	305	16	T20075	Human gene express
c	18	215	6.1	300	20	Z17757	Human gene express
c	19	215	6.1	300	20	Z14226	Human gene express
c	20	210	6.0	767	20	Z16120	Human gene express
c	21	55	1.6	1000	21	A02484	Human colon cancer
c	22	52	1.5	12001	16	Q76213	HSV L/ST region.
c	23	51.8	1.5	2188	20	Z77506	Human ovarian tumo
c	24	51.6	1.5	936	20	X56375	Human DNA-dependen
c	25	51.2	1.5	53526	19	T94101	Human PKD1 gene.
c	26	51.2	1.5	53577	17	T18551	Human polycystic k
c	27	51.2	1.5	53577	19	T94108	Human PKD1 locus b
c	28	50.4	1.4	10732	21	A10594	Gene encoding a su
c	29	49	1.4	1337	20	Z17263	Human gene express
c	30	48.8	1.4	867	21	Z98056	Human secreted pro
c	31	48.6	1.4	15899	19	V62164	HSV-2 strain SB5 C
c	32	48.6	1.4	16812	19	V62175	HSV-2 strain SB5 C
c	33	48.4	1.4	38186	20	Z32028	Human METH1 relate
c	34	48	1.4	8670	20	Z32059	Human METH2 relate
c	35	47.4	1.4	3198	20	X02974	Human IL-lra BAC c
c	36	47	1.3	1235	20	Z16147	Human gene express
c	37	47	1.3	23142	21	A35158	Human gene express
c	38	47	1.3	30417	21	A35160	Human adenosine re
c	39	45.8	1.3	1446	20	Z16805	Human adenosine re
c	40	45.4	1.3	80240	20	V83940	Human gene express
c	41	45.4	1.3	80595	20	V83939	NC-contlig derived
c	42	45	1.3	4897	11	Q03259	HC-contlig derived
c	43	44.8	1.3	1102	20	Z16923	Pseudorabies virus
c	44	44.8	1.3	1725	12	Q11243	Human gene express
c	45	44.6	1.3	1558	12	Q10896	PRV 28k gene. pse
							Encodes Xenopus Bo

ALIGNMENTS

RESULT 1	
ID Z933322	standard; cDNA; 3506 BP.
XX AC Z933322;	
XX DT 04-JUL-2000	(first entry)
XX DE Human stauflen cDNA.	
XX XX	
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW acquired immune deficiency syndrome; protease; human; ss.	
XX OS Homo sapiens.	
XX FH Key	Location/Qualifiers
FT CDS	409..2142
FT FT	/*tag= a
FT FT	/product= Stauflen
XX PN CA2238656-A1.	
XX PD 22-NOV-1999.	
XX PF 22-MAY-1998;	98CA-2238656.
XX PR 22-MAY-1998;	98CA-2238656.
XX PA (UYMO-) UNIV MONTREAL.	
XX XX	
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;	
PI Moulard AJ;	
XX DR WPI; 2000-246924/22.	

DR P-PSDB; Y83023.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

Query Match 99.9%; Score 3502.8; DB 21; Length 3506;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3504; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 acttcctgcccggctgcggcgccctgagcgctcttcagcgtttgcggcgcgctgcgct 60
DB 1 acttcctgcccggcgcgcgccctgagcgctcttcagcgtttgcggcgcgctgcgct 60
QY 61 ctctctcgctcccgctctctttgacgcgctcccccgcggcgcgcgccgctc 120
DB 61 ctctctcgctcccgctctctttgacgcgctcccccgcggcgcgcgccgctc 120
QY 121 ctccacggcgaactccgcctctctccctccttcgctcctctcctccttttccct 180
DB 121 ctccacggcgaactccgcctctctccctccttcgctcctctcctccttttccct 180
QY 181 ctctcctccctcctcgcgcgcacccgcgcaggaacgcggcggggagcagctcgagc 240
DB 181 ctctcctccctcctcgcgcgcacccgcgcaggaacgcggcggggagcagctcgagc 240
QY 241 agcagccagagttattaaaccacttaacctcagaactgaaacaagaacattgtcc 300
DB 241 agcagccagagttattaaaccacttaacctcagaactgaaacaagaacattgtcc 300
QY 301 tggagcgccctcttttaaaaaagtagaactttagacttcatagcactgaattaactg 360
DB 301 tggagcgccctcttttaaaaaagtagaactttagacttcatagcactgaattaactg 360
QY 361 cactgaaagctgttttaacctgtcatttgttcaactttgttgaagtgaccatgtctcaagt 420
DB 361 cactgaaagctgttttaacctgtcatttgttcaactttgttgaagtgaccatgtctcaagt 420
QY 421 caagtgcaagttcagaaaccacatctgtctctcctcagggagccaataactgaaagaac 480
DB 421 caagtgcaagttcagaaaccacatctgtctctcctcagggagccaataactgaaagaac 480
QY 481 cagtctctctcaccagcctttagatgatatccttactactagctctcgcctct 540
DB 481 cagtctctctcaccagcctttagatgatatccttactactagctctcgcctct 540
QY 541 gaaatgcaggtagaccattcaaaactcgtttaccctctgcacatctatataccacc 600
DB 541 gaaatgcaggtagaccattcaaaactcgtttaccctctgcacatctatataccacc 600
QY 601 agtgcagctgcagaaagcataaccctactgtagaactaaatgcactgtgtgaactc 660
DB 601 agtgcagctgcagaaagcataaccctactgtagaactaaatgcactgtgtgaactc 660

DB 601 agtgcagctgcagaaagcataaccctactgtagaactaaatgcactgtgtgaactc 660
QY 661 ggaaaaaaaccaatgataagcctgttgaaccttactctcgatgcagtcaccataaac 720
DB 661 ggaaaaaaaccaatgataagcctgttgaaccttactctcgatgcagtcaccataaac 720
QY 721 tacaacatgagagagtgcttatcccccaggttactttaccatttccagttccact 780
DB 721 tacaacatgagagagtgcttatcccccaggttactttaccatttccagttccact 780
QY 781 ttactttatcaagtggaaacttctctgtggaaggacagcaatttaatggcaagaagaca 840
DB 781 ttactttatcaagtggaaacttctctgtggaaggacagcaatttaatggcaagaagaca 840
QY 841 agacagctgcgaaacacagatcgtcgtcgcacaaagcgttgagatcctgcagaatgagcc 900
DB 841 agacagctgcgaaacacagatcgtcgtcgcacaaagcgttgagatcctgcagaatgagcc 900
QY 901 ctgcagagagctggaagtgaatgtgaagagaatccgaagaagaaatctcaataatct 960
DB 901 ctgcagagagctggaagtgaatgtgaagagaatccgaagaagaaatctcaataatct 960
QY 961 gaataagtcagaagtgttagatgtgcacttaaacggaacttgctgtgaatttcgagtg 1020
DB 961 gaataagtcagaagtgttagatgtgcacttaaacggaacttgctgtgaatttcgagtg 1020
QY 1021 gcccggaagatggcccaaccacatgaagaactttgtgaccaaggttcggttgaggag 1080
DB 1021 gcccggaagatggcccaaccacatgaagaactttgtgaccaaggttcggttgaggag 1080
QY 1081 ttgtggyggaagtgaagggaagaagcaagaagatttcaagaagaaatgcgcacatagct 1140
DB 1081 ttgtggyggaagtgaagggaagaagcaagaagatttcaagaagaaatgcgcacatagct 1140
QY 1141 gtctctgagagctggaagaagttlacccgcccctgcctgcagttgaacgagtaagcctaga 1200
DB 1141 gtctctgagagctggaagaagttlacccgcccctgcctgcagttgaacgagtaagcctaga 1200
QY 1201 atcaaaaagaaacaaacccatagttcaagccacagacagcaagccagaalatgagccagg 1260
DB 1201 atcaaaaagaaacaaacccatagttcaagccacagacagcaagccagaalatgagccagg 1260
QY 1261 atcaatccgattagccgaactggccagatccagacagcgaagaaagagagccagag 1320
DB 1261 atcaatccgattagccgaactggccagatccagacagcgaagaaagagagccagag 1320
QY 1321 tacacgctcctcacagagcgagcgctcccgcgccgagggagttgtgatgcaggtgaag 1380
DB 1321 tacacgctcctcacagagcgagcgctcccgcgccgagggagttgtgatgcaggtgaag 1380
QY 1381 gttggaaccacacactgcagagaaggaacggygcaccaacaagaagtyggccaagcgcaatgca 1440
DB 1381 gttggaaccacacactgcagagaaggaacggygcaccaacaagaagtyggccaagcgcaatgca 1440
QY 1381 gttggaaccacacactgcagagaaggaacggygcaccaacaagaagtyggccaagcgcaatgca 1440
DB 1381 gttggaaccacacactgcagagaaggaacggygcaccaacaagaagtyggccaagcgcaatgca 1440
QY 1441 gccgagaacatgctggaagatccttggttcaaaagtcgccgagcgcgagcccaaaccc 1500
DB 1441 gccgagaacatgctggaagatccttggttcaaaagtcgccgagcgcgagcccaaaccc 1500
QY 1501 gcaactcaagtccagaggaagacacccataaagaacacaggggagtgaagaaagtaacc 1560
DB 1501 gcaactcaagtccagaggaagacacccataaagaacacaggggagtgaagaaagtaacc 1560
QY 1561 tttttgaacctggtctgtgggagtgaagaaatgggactagtaataaagagatgattcag 1620
DB 1561 tttttgaacctggtctgtgggagtgaagaaatgggactagtaataaagagatgattcag 1620
QY 1621 atgccttatctaagtcatacagcagctgcctgtggaattcttcccatgtgtgccgaggtc 1680
DB 1621 atgccttatctaagtcatacagcagctgcctgtggaattcttcccatgtgtgccgaggtc 1680
QY 1681 gcccaggtgtgtagagttagtcagaagacatcacacaaagatttaccagggcagctccg 1740
DB 1681 gcccaggtgtgtagagttagtcagaagacatcacacaaagatttaccagggcagctccg 1740

QY 1741 aatcctgccaagggccaagcgttaactgcatgatagcccgagagttgtgtatg9gggacc 1800
Db 1741 aatcctgccaagggccaagcgttaactgcatgatagcccgagagttgtgtatg9gggacc 1800
QY 1801 tcgcccacagcccgagaccattttaagaataacatctcttcagggccaagcgtaccccatgga 1860
Db 1801 tcgcccacagcccgagaccattttaagaataacatctcttcagggccaagcgtaccccatgga 1860
QY 1861 cctctcacgagaccctctgtagcgaactggaactatcttccagagtcagggatccaggtt 1920
Db 1861 cctctcacgagaccctctgtagcgaactggaactatcttccagagtcagggatccaggtt 1920
QY 1921 gaatacaaaagacttccccaaaaacaagaagaaattgtatctctatcaattgctcc 1980
Db 1921 gaatacaaaagacttccccaaaaacaagaagaaattgtatctctatcaattgctcc 1980
QY 1981 tctcagccacctctgtagcagccatggtatcggaagatgttgagtcctgcatgatg 2040
Db 1981 tctcagccacctctgtagcagccatggtatcggaagatgttgagtcctgcatgatg 2040
QY 2041 gctgcgctggaacatcttaagtgtgctgtgagttggaccaacaagtaacagatgcca 2100
Db 2041 gctgcgctggaacatcttaagtgtgctgtgagttggaccaacaagtaacagatgcca 2100
QY 2101 agaacaggaagaaacggaaccaatgtctgtgtggtggaggtgctgaacctttctgycatgaa 2160
Db 2101 agaacaggaagaaacggaaccaatgtctgtgtggtggaggtgctgaacctttctgycatgaa 2160
QY 2161 ccattataaaaaccccaaacataataactgaaataactgaaactgcttggaaaaattggaat 2220
Db 2161 ccattataaaaaccccaaacataataactgaaataactgaaataactgcttggaaaaattggaat 2220
QY 2221 ttctgatacctccagtggtggccggaagagacaagtggtgtaaaagatgtggcagcagcaggg 2280
Db 2221 ttctgatacctccagtggtggccggaagagacaagtggtgtaaaagatgtggcagcagcaggg 2280
QY 2281 aagacaacagaaacacacaagaggcggtgtggtggcggctggaactgtgctggggtttgtgt 2340
Db 2281 aagacaacagaaacacacaagaggcggtgtggtggcggctggaactgtgctggggtttgtgt 2340
QY 2341 gatggccaactcggtgaacctggtgtccctacgcaatagcagctgctgtggtggaaagaag 2400
Db 2341 gatggccaactcggtgaacctggtgtccctacgcaatagcagctgctgtggtggaaagaag 2400
QY 2401 gctgccagcagcagctgtgtctcccgggacaccagcagatccacaccctgggcacctcgt 2460
Db 2401 gctgccagcagcagctgtgtctcccgggacaccagcagatccacaccctgggcacctcgt 2460
QY 2461 gtttgctcttttttccctgtgtgaagaagaagaacggaacggaaccccttccaagctgg 2520
Db 2461 gtttgctcttttttccctgtgtgaagaagaagaacggaacggaaccccttccaagctgg 2520
QY 2521 ctcaactcagacacatattgggacaaaaccctggacagccaatggcagaagagcccttggaccg 2580
Db 2521 ctcaactcagacacatattgggacaaaaccctggacagccaatggcagaagagcccttggaccg 2580
QY 2581 gccccaagagctaaaaagcaccagagaaaaatcaaatgcttccctactcagcgtgaccaactt 2640
Db 2581 gccccaagagctaaaaagcaccagagaaaaatcaaatgcttccctactcagcgtgaccaactt 2640
QY 2641 ttctagtgtagcagggccccaacccctcctgtagtaccacacacacatcaccacgttctct 2700
Db 2641 ttctagtgtagcagggccccaacccctcctgtagtaccacacacacatcaccacgttctct 2700
QY 2701 ctccaacagtgatctgtatcttagttcatatttcttcttgatgtgatagacactat 2760
Db 2701 ctccaacagtgatctgtatcttagttcatatttcttcttgatgtgatagacactat 2760
QY 2761 ataaaatttcatatttgagaatttctcaattgtatcttagttaataatagcacagtttggaac 2820
Db 2761 ataaaatttcatatttgagaatttctcaattgtatcttagttaataatagcacagtttggaac 2820

QY 2821 ttgtctgagactgactttatcaataatctaacccgacaagaatcatatccatgtgtatgtg 2880
Db 2821 ttgtctgagactgactttatcaataatctaacccgacaagaatcatatccatgtgtatgtg 2880
QY 2881 gtagaacattttatctcatgtaactaacccagagacagtttcagtgatgcaaatgtgtgc 2940
Db 2881 gtagaacattttatctcatgtaactaacccagagacagtttcagtgatgcaaatgtgtgc 2940
QY 2941 cctctgttcagctgaaacagtcctggaacttcaaaaacctgtaataagtcctccacagt 3000
Db 2941 cctctgttcagctgaaacagtcctggaacttcaaaaacctgtaataagtcctccacagt 3000
QY 3001 tgtataaattggaacaatttaggaattttaactttagatgatacttggttccatttta 3060
Db 3001 tgtataaattggaacaatttaggaattttaactttagatgatacttggttccatttta 3060
QY 3061 ttcatattttattttgtttaatgcaaacaggaacttaaatgaactttgatactctgttttaa 3120
Db 3061 ttcatattttattttgtttaatgcaaacaggaacttaaatgaactttgatactctgttttaa 3120
QY 3121 agattataaaaaaacattgtgtatctatatacatatggtccttgaggacttagcttcaacta 3180
Db 3121 agattataaaaaaacattgtgtatctatatacatatggtccttgaggacttagcttcaacta 3180
QY 3181 cactacagatatgatctccatgtagtccatatataaaacctgcagagtgatatttccagagtg 3240
Db 3181 cactacagatatgatctccatgtagtccatatataaaacctgcagagtgatatttccagagtg 3240
QY 3241 ctcgatacgtttaattacatctccatttagggctgnaaaagaatgacctagcttccgtata 3300
Db 3241 ctcgatacgtttaattacatctccatttagggctgnaaaagaatgacctagcttccgtata 3300
QY 3301 cagctgtgtgtcttcttgatgtgtgtgtactgttacacagaagtgtgtgcactgagctctg 3360
Db 3301 cagctgtgtgtcttcttgatgtgtgtgtactgttacacagaagtgtgtgcactgagctctg 3360
QY 3361 cgtgtgtccgtatgnaaaacctggtagccctgcgagttgaagtaactgttccatccaatg 3420
Db 3361 cgtgtgtccgtatgnaaaacctggtagccctgcgagttgaagtaactgttccatccaatg 3420
QY 3421 tttaacgtggaatttttccccaatggaatgtaagtaaaacttaagtgttgcatacaat 3480
Db 3421 tttaacgtggaatttttccccaatggaatgtaagtaaaacttaagtgttgcatacaat 3480
QY 3481 aaatgtaataactaaaaaanaaaa 3506
Db 3481 aaatgtaataactaaaaaanaaaa 3506

RESULT 2
293323
ID Z93323 standard; cDNA; 3217 BP.
XX
AC Z93323;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 363..1853
FT /tag= a
FT /product= Stauflen
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.

XX 22-MAY-1998; 98CA-2238656.
PF
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroseillers L;
PI Moulard AJ;
XX
DR WPI: 2000-246924/22.
DR P-PSDB; Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1a; 96pp; English.
XX
CC Stauflen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC stauflen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian stauflen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauflen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC Z93322.
XX
SQ Sequence 3217 BP; 889 A; 796 C; 742 G; 790 T; 0 other;

Query Match 83.1%; Score 2914.8; DB 21; Length 3217;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 3215; Conservative 0; Mismatches 2; Indels 289; Gaps 1;

QY 1 actctcgccggtcgcggtcgccgtgagcgctcttcagcggttgcgcggtcgcggt 60
|||
Db 1 actctcgccggtcgcggtcgccgtgagcgctcttcagcggttgcgcggtcgcggt 60
QY 61 ctctctcggtccggtctcttgaccgctcccccccgccggtcgcggtcgccgtc 120
|||
Db 61 ctctctcggtccggtctcttgaccgctcccccccgccggtcgcggtcgccgtc 120
QY 121 ctccacggtccggtctcttcctccctcgctcccttctctctctctctctctt 180
|||
Db 121 ctccacggtccggtctcttcctccctcgctcccttctctctctctctctctt 180
QY 181 ctctctccctctcgccgacccggtcagcggtcggtcggtcggtcggtcggtc 240
|||
Db 181 ctctctccctctcgccgacccggtcagcggtcggtcggtcggtcggtcggtc 240
QY 241 agcagcagagtttataaccacttaacctcagaaactgaacaaagacaacattgtcc 300
|||
Db 241 agcagcagagtttataaccacttaacctcagaaactgaacaaagacaacattgtcc 300
QY 301 tggagcgccctcttttaaaaaagtagaactttagacttcataagcactgaattaacctg 360
|||
Db 301 tggagcgccctcttttaaaaaagtagaactttagacttcataagcactgaattaacctg 360
QY 361 cactgaaagctgttctacactgtcatttgttcaactttgttgaaagtgcacatgtctcaagt 420
323 ----- 322
QY 421 caagtgcaggtcagaaccacatctgtctctctcagggagccaataactgaacaagaac 480

Db 323 ----- 322
QY 481 cagtcctctctcaacgaccttgatgattcttctactactagctctctgacctc 540
Db 323 ----- 322
QY 541 gaaatgcagtagaacccattcaaaaactctgttaccctctgcattctataccacc 600
Db 323 ----- 322
QY 601 agtgcagctgcagaaagcataaccctactgtagaactaaatgcactgtgcatgaaact 660
|||
Db 323 -----agaaagcataaccctactgtagaactaaatgcactgtgcatgaaact 371
QY 661 gaaaaaaaccaaactgtataagcctgttgaccctactctcgatgcagtcacactataac 720
|||
Db 372 gaaaaaaaccaaactgtataagcctgttgaccctactctcgatgcagtcacactataac 431
QY 721 tacacatgagaggtgtcttatcccgaggtactttaccacattccagttccacct 780
|||
Db 432 tacacatgagaggtgtcttatcccgaggtactttaccacattccagttccacct 491
QY 781 ttacttatcaagtgaacttctgtgtgagggacagcaatttaatgycaaagaaagaca 840
|||
Db 492 ttacttatcaagtgaacttctgtgtgagggacagcaatttaatgycaaagaaagaca 551
QY 841 agacaggtcgcgaacaacgatgtctgtccaaagcgttgagatcctgcagaatgagcc 900
|||
Db 552 agacaggtcgcgaacaacgatgtctgtccaaagcgttgagatcctgcagaatgagcc 611
QY 901 ctgcagagaggtgtgaggtgtaagtgaagaaatccgaagaagaaatctcaataatct 960
|||
Db 612 ctgcagagaggtgtgaggtgtaagtgaagaaatccgaagaagaaatctcaataatct 671
QY 961 gaataagtcaagtgtttagatgtgcaacttaacggaactgtcgtgtaatttcgaggtg 1020
|||
Db 672 gaataagtcaagtgtttagatgtgcaacttaacggaactgtcgtgtaatttcgaggtg 731
QY 1021 gcccggtgaggtgtgccccacacatgaagaacttgtgaccaaaggttcgtgtgggag 1080
|||
Db 732 gcccggtgaggtgtgccccacacatgaagaacttgtgaccaaaggttcgtgtgggag 791
QY 1081 ttgtgtggggaagtgtgaaggaaagcaagaagtcttcaaaagaaatgtccgcatagtc 1140
|||
Db 792 ttgtgtggggaagtgtgaaggaaagcaagaagtcttcaaaagaaatgtccgcatagtc 851
QY 1141 gtctctgagaggtgaagaagtltaacgccccctgtcgtcagtltaacgagtltaagcctga 1200
|||
Db 852 gtctctgagaggtgaagaagtltaacgccccctgtcgtcagtltaacgagtltaagcctga 911
QY 1201 atcaaaaagaaaacaaacccatagtcacagccacagacaagccagaatatgtgccaggg 1260
|||
Db 912 atcaaaaagaaaacaaacccatagtcacagccacagacaagccagaatatgtgccaggg 971
QY 1261 atcaatccgatttagccgactgtgccagatccagcagcagcaaaaaggaagagccagag 1320
|||
Db 972 atcaatccgatttagccgactgtgccagatccagcagcagcaaaaaggaagagccagag 1031
QY 1321 tacacgtctctcaacagagcgaggtctcccgccgagcgaggtgtgtgatgtcaggtgaag 1380
|||
Db 1032 tacacgtctctcaacagagcgaggtctcccgccgagcgaggtgtgtgatgtcaggtgaag 1091
QY 1381 gtggaaaacacactgcagaaggaagcagccaaacaagaaggtggtccaaagcgcaatgca 1440
|||
Db 1092 gtggaaaacacactgcagaaggaagcagccaaacaagaaggtggtccaaagcgcaatgca 1151
QY 1441 gccgagaacatgtgtgagatcctgtgttcaaaagtcgccgagcgagcccaaccaaaccc 1500
|||
Db 1152 gccgagaacatgtgtgagatcctgtgttcaaaagtcgccgagcgagcccaaccaaaccc 1211
QY 1501 gcactcaagtgcaggtgagaagacacccataaagaacacaggggtatggaagaaagtaacc 1560
|||

Db 1212 gcactcaagtcagagagagaacacccataaagaaaccagggatgtgaagaaaagtaacc 1271
QY 1561 tttttgaacctggtctctggtgatgtaaaaatggtgactagtaataaagagatgattcagg 1620
Db 1272 tttttgaacctggtctctggtgatgtaaaaatggtgactagtaataaagagatgattcagg 1331
QY 1621 atgccttatctaagtcatacagcagctgctgtgctggaattctcccatgtgtcccgaggtc 1680
Db 1332 atgccttatctaagtcatacagcagctgctgtgctggaattctcccatgtgtcccgaggtc 1391
QY 1681 gcccaagctgtagagatgaagacatcacaccaaaagattttaccagggcagctccg 1740
Db 1392 gcccaagctgtagagatgaagacatcacaccaaaagattttaccagggcagctccg 1451
QY 1741 aatcctgccaaagggccaggttaactgcatgtagcccgagagtgtgtatgtggggcacc 1800
Db 1452 aatcctgccaaagggccaggttaactgcatgtagcccgagagtgtgtatgtggggcacc 1511
QY 1801 tcgcccacagccgagaccattttaagaataacatctcttcaggccacgttaaccatgga 1860
Db 1512 tcgcccacagccgagaccattttaagaataacatctcttcaggccacgttaaccatgga 1571
QY 1861 cctctcacgagaccctctgtagcaactgtgactatcttcagaggtccaggatccaggtt 1920
Db 1572 cctctcacgagaccctctgtagcaactgtgactatcttcagaggtccaggatccaggtt 1631
QY 1921 gaatacaaaagacttccccaaaaacaacaagaacgaatttgtatctcttaacatgtctcc 1980
Db 1632 gaatacaaaagacttccccaaaaacaacaagaacgaatttgtatctcttaacatgtctcc 1691
QY 1981 tctcagccacctctgtatcagccatgtgtatcggcaaggaatgtgagtccttgccatgatg 2040
Db 1692 tctcagccacctctgtatcagccatgtgtatcggcaaggaatgtgagtccttgccatgatg 1751
QY 2041 gctgcgctgtaacatcttaaaagtgtgtctgtcgtgagttggaaccaaaagtatacagagatgcc 2100
Db 1752 gctgcgctgtaacatcttaaaagtgtgtctgtcgtgagttggaaccaaaagtatacagagatgcc 1811
QY 2101 agaacaaggaaacggaaccaatgtctgtgtggtggaggtgtctgaaacctttctggtccatgaa 2160
Db 1812 agaacaaggaaacggaaccaatgtctgtgtggtggaggtgtctgaaacctttctggtccatgaa 1871
QY 2161 ccattataaaatcccaacatatatactgnaaaatactgnaaactgctttgnaaaatttggat 2220
Db 1872 ccattataaaatcccaacatatatactgnaaaatactgnaaactgctttgnaaaatttggat 1931
QY 2221 ttctgatacctccagtggtggtccgagagacacacggtgtggttaaaggatgtgtggcagcagcagg 2280
Db 1932 ttctgatacctccagtggtggtccgagagacacacggtgtggttaaaggatgtgtggcagcagcagg 1991
QY 2281 aagaacaacagaaacaacaaggaaggcggctgtgtgcccggctggaactgtgtggtgggtttgtgt 2340
Db 1992 aagaacaacagaaacaacaaggaaggcggctgtgtgcccggctggaactgtgtggtgggtttgtgt 2051
QY 2341 gatggccaactcgtgtgacctggtgtccctacgcaatlagcagctgcctgtggtgggaagaagg 2400
Db 2052 gatggccaactcgtgtgacctggtgtccctacgcaatlagcagctgcctgtggtgggaagaagg 2111
QY 2401 gctgcccagccagctgtgtctctcccggtgacaccagcagatccacaccctgtggcaccctcgt 2460
Db 2112 gctgcccagccagctgtgtctctcccggtgacaccagcagatccacaccctgtggcaccctcgt 2171
QY 2461 gtttggtctttttttccctgtgtgaagaagaacacggcacgaccccttctcaagtgtg 2520
Db 2172 gtttggtctttttttccctgtgtgaagaagaacacggcacgaccccttctcaagtgtg 2231
QY 2521 ctcaactcagacacattgtggacaaacccctgtgacagcccatgtccagagagagcctttgaccg 2580
Db 2232 ctcaactcagacacattgtggacaaacccctgtgacagcccatgtccagagagagcctttgaccg 2291
QY 2581 gccccaagagctaaaaagcacaagagaaaaatcaaatgtcttctactcagcgtgaaccaactt 2640
Db 2292 gccccaagagctaaaaagcacaagagaaaaatcaaatgtcttctactcagcgtgaaccaactt 2351

QY 2641 ttctagtgtgccacggccccaccacctctcgtcagtlacccacacatcaccaactgttctc 2700
Db 2352 ttctagtgtgccacggccccaccacctctcgtcagtlacccacacatcaccaactgttctc 2411
QY 2701 ctccaacagtgatctgtattctcttagtttcatatttctctttgtattgatatgacactat 2760
Db 2412 ctccaacagtgatctgtattctcttagtttcatatttctctttgtattgatatgacactat 2471
QY 2761 ataaaatttcatatttgagaatttctcaattgttatcttaagtaaatagcacagtttggaaac 2820
Db 2472 ataaaatttcatatttgagaatttctcaattgttatcttaagtaaatagcacagtttggaaac 2531
QY 2821 ttgtctgagactgactttatcataaaltctaacggacaaagatcatatccatgtgtatgtg 2880
Db 2532 ttgtctgagactgactttatcataaaltctaacggacaaagatcatatccatgtgtatgtg 2591
QY 2881 gtagacatttttatttcatgtactaacccagagacagtttcagtgatgcaaatgtgtgc 2940
Db 2592 gtagacatttttatttcatgtactaacccagagacagtttcagtgatgcaaatgtgtgc 2651
QY 2941 cctctgttcagctgtaaacaggtcctgtgacttttcaaaaaaccttgataagtctccacagt 3000
Db 2652 cctctgttcagctgtaaacaggtcctgtgacttttcaaaaaaccttgataagtctccacagt 2711
QY 3001 tgtataaatttgacaatttagaattttaaactttagatgacatattgtgttccattttta 3060
Db 2712 tgtataaatttgacaatttagaattttaaactttagatgacatattgtgttccattttta 2771
QY 3061 ttcatatttatttttgtttaatgcaaaacagagacttaaatgaaactttgactctgttttaa 3120
Db 2772 ttcatatttatttttgtttaatgcaaaacagagacttaaatgaaactttgactctgttttaa 2831
QY 3121 agattatataaaaaacatgtgtatctatatacatatggtccttgagagacttagcttcaacta 3180
Db 2832 agattatataaaaaacatgtgtatctatatacatatggtccttgagagacttagcttcaacta 2891
QY 3181 cactacaggatatgatctccatgtagtcacatataaaccctgcagagtgtatttccagagtg 3240
Db 2892 cactacaggatatgatctccatgtagtcacatataaaccctgcagagtgtatttccagagtg 2951
QY 3241 ctgcatactgttaattacatctccattagggctgaaaagaatgacctagcttctgtata 3300
Db 2952 ctgcatactgttaattacatctccattagggctgaaaagaatgacctagcttctgtata 3011
QY 3301 cagctgtgtgcttttgatgtgtgttactgtacacagaagtgtgtgcactgaggtctgtg 3360
Db 3012 cagctgtgtgcttttgatgtgtgttactgtacacagaagtgtgtgcactgaggtctgtg 3071
QY 3361 cgtgtgtccglatgnaaacctgtgtagccctgcgagtttaagtaactgtctccatcatgtg 3420
Db 3072 cgtgtgtccglatgnaaacctgtgtagccctgcgagtttaagtaactgtctccatcatgtg 3131
QY 3421 tttaacgttggaatttttctcccatggaatgttaagtaaaaacttaagtgtttgtcatcaat 3480
Db 3132 tttaacgttggaatttttctcccatggaatgttaagtaaaaacttaagtgtttgtcatcaat 3191
QY 3481 aaatggtataactataaaaaaa 3506
Db 3192 aaatggtataactataaaaaaa 3217

RESULT 3
Z93324
ID Z93324 standard; cDNA; 3142 BP.
XX
AC Z93324;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;

KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 288..1193
FT /*tag= a
FT /product= staufen protein
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI Moulard AJ;
XX
DR WPI; 2000-246924/22.
DR P-PSDB; Y83108.
XX
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
PS
PS Claim 4; Figure 1b; 96pp; English.
XX
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
XX
SQ Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 82.6%; Score 2896; DB 21; Length 3142;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 611 cagaaagcataaccctactgtagaactaatgacctgtgcatgaaacttgaaaaaac 670
Db 247 cagaagcataaccctactgtagaactaatgacctgtgcatgaaacttgaaaaaac 306
QY 671 caatgtataagcctgtgaccttactctcgatgagtcacactataactacaatga 730
Db 307 caatgtataagcctgtgaccttactctcgatgagtcacactataactacaatga 366
QY 731 gaggagtgctatcccccgaggtactttaccatttccagttccactttacttattc 790
Db 367 gaggagtgctatcccccgaggtactttaccatttccagttccactttacttattc 426
QY 791 aagtgaacttctgtgagagacagcaattaatgcaaaagaaagacaagacagctg 850
Db 427 aagtgaacttctgtgagagacagcaattaatgcaaaagaaagacaagacagctg 486

QY 851 cgaaaacacgatgtctgtccaaaagcglttgagatccttcgagaatgaagccctgcagaga 910
Db 487 cgaaacacgatgtctgtccaaaagcglttgagatccttcgagaatgaagccctgcagaga 546
QY 911 ggtgaggtgaatggaagagaatccgaagaagaaaatctcaataatctgaataagtc 970
Db 547 ggtgaggtgaatggaagagaatccgaagaagaaaatctcaataatctgaataagtc 606
QY 971 aagtgttgagattgcaacttaaacggaactgtcgttgaaatttcgaggtgccccggaga 1030
Db 607 aagtgttgagattgcaacttaaacggaactgtcgttgaaatttcgaggtgccccggaga 666
QY 1031 gtgccccaccacatgaagaacttltgtgaccgaagtttcggttgagggttgttgagg 1090
Db 667 gtgccccaccacatgaagaacttltgtgaccgaagtttcggttgagggttgttgagg 726
QY 1091 aaggtgaagggaagcaagaagaatttcaagaagaaaatgccccatagctgttcttgagg 1150
Db 727 aaggtgaagggaagcaagaagaatttcaagaagaaaatgccccatagctgttcttgagg 786
QY 1151 agctgaagaagttaccgccccctgtcctgtcagttgaaagtaagcagtaaatcaaaaaga 1210
Db 787 agctgaagaagttaccgccccctgtcctgtcagttgaaagtaagcagtaaatcaaaaaga 846
QY 1211 aaacaaaccatagtccaagccacagacagacagccagaaatattgccccggatcaatccga 1270
Db 847 aaacaaaccatagtccaagccacagacagacagccagaaatattgccccggatcaatccga 906
QY 1271 ttagccgactggtcccaagatccagcagcgaataaaggagaaggagccagagtacacgtcc 1330
Db 907 ttagccgactggtcccaagatccagcagcgaataaaggagaaggagccagagtacacgtcc 966
QY 1331 tcacagagcgaagcctcccgccgcccagcagggagtttgtatgcaggtgaaggttggaaacc 1390
Db 967 tcacagagcgaagcctcccgccgcccagcagggagtttgtatgcaggtgaaggttggaaacc 1026
QY 1391 acaactgagaagaagcgggcaccaacaagaagttgccaagcgcaatgcagccgagaaca 1450
Db 1027 acaactgagaagaagcgggcaccaacaagaagttgccaagcgcaatgcagccgagaaca 1086
QY 1451 tgcctgagatccttgttccaagttcccgagcggcagcccaaccacccgactcaagt 1510
Db 1087 tgcctgagatccttgttccaagttcccgagcggcagcccaaccacccgactcaagt 1146
QY 1511 cagaggaagacacccataaagaacccagggagtgaagaagaagttaaccttittgaac 1570
Db 1147 cagaggaagacacccataaagaacccagggagtgaagaagaagttaaccttittgaac 1206
QY 1571 ctggtcttgggatatgaatgtgactagtaataaagagatgagttcagatgccttattc 1630
Db 1207 ctggtcttgggatatgaatgtgactagtaataaagagatgagttcagatgccttattc 1266
QY 1631 taagtcatcagcagctgtcgtgtgaaattcttcccatgtgtgccgaaggtcgccagctg 1690
Db 1267 taagtcatcagcagctgtcgtgtgaaattcttcccatgtgtgtccgaaggtcgccagctg 1326
QY 1691 taggagtagtcaaggaacatcacaccaagaatttaccagggcagctccgaatcctgcca 1750
Db 1327 taggagtagtcaaggaacatcacaccaagaatttaccagggcagctccgaatcctgcca 1386
QY 1751 aggcacggtactgcgatgatagcccgagagttgtgtatggyggcacctcgccacag 1810
Db 1387 aggcacggtactgcgatgatagcccgagagttgtgtatggyggcacctcgccacag 1446
QY 1811 ccgagacattttaagaataaacatctctcagggccaagtaaccatggaactctcagga 1870
Db 1447 ccgagacattttaagaataaacatctctcagggccaagtaaccatggaactctcagga 1506
QY 1871 gaccctctgagcaactggactatcttccagagttccagggattccaggttgaatacaag 1930
Db 1507 gaccctctgagcaactggactatcttccagagttccagggattccaggttgaatacaag 1566
QY 1931 acttccccaaaacaagaagaatltgtatctcttatcaattgtctctcagccac 1990

Db 1567 actccccaaaaacaacaagaacgaatttgtatctcttatcaattgtctccctcagccac 1626
QY 1991 cctctgatacgccatgtgtatccgcaagatgtgtgagtcctgcatagtatgctgcgtga 2050
Db 1627 ctctgatcagccatggtatccgcaagatgtgtgagtcctgcatagtatgctgcgtga 1686
QY 2051 acatcttaaaagtgtctgtgaglttgaccaaaagtacagagatgccaagaacaggaa 2110
Db 1687 acatcttaaaagtgtctgtgaglttgaccaaaagtacagagatgccaagaacaggaa 1746
QY 2111 acggaaccaatgtctgtgtggaaggtgtgtaacctttctgagccatggaaccattataa 2170
Db 1747 acggaaccaatgtctgtgtggaaggtgtgtaacctttctgagccatggaaccattataa 1806
QY 2171 atcccaacatatatactgaaataactgtaactgtttgaaatttgaaatttctgtatacc 2230
Db 1807 atcccaacatatatactgaaataactgtaactgtttgaaatttctgtatacc 1866
QY 2231 tccagttggccgagagacacacggttggtaaaagatgtggcgagcagcagggaagaacaag 2290
Db 1867 tccagttggccgagagacacacggttggtaaaagatgtggcgagcagcagggaagaacaag 1926
QY 2291 aaaaacaagaagcggtgtggtccgaacatagcagctgtctgtgggaagaagcgctccagc 2350
Db 1927 aaaaacaagaagcggtgtggtccgaacatagcagctgtctgtgggaagaagcgctccagc 1986
QY 2351 cgtgtgacctgtgcgttccctacgaacatagcagctgtctgtgggaagaagcgctccagc 2410
Db 1987 cgtgtgacctgtgcgttccctacgaacatagcagctgtctgtgggaagaagcgctccagc 2046
QY 2411 cagctggttctcccgagacacacagcagatccacacctgggacactccgtgttgcctt 2470
Db 2047 cagctggttctcccgagacacacagcagatccacacctgggacactccgtgttgcctt 2106
QY 2471 tttttccctgtgtgaagaagaacgcgcagcacccttctcaagctgtgtactcaga 2530
Db 2107 tttttccctgtgtgaagaagaacgcgcagcacccttctcaagctgtgtactcaga 2166
QY 2531 cacattgggacaaaacctgtgacagccatgcccagagagagccttgaaccggcccccagagc 2590
Db 2167 cacattgggacaaaacctgtgacagccatgcccagagagagccttgaaccggcccccagagc 2226
QY 2591 taaaagcaccagagaaaataatgcttctactcagcgltgacccaaacttttctagtgtg 2650
Db 2227 taaaagcaccagagaaaataatgcttctactcagcgltgacccaaacttttctagtgtg 2286
QY 2651 ccacggcccccacacacctctgtcagtaaccacacatcaccaactgttctcttccaacag 2710
Db 2287 ccacggcccccacacacctctgtcagtaaccacacatcaccaactgttctcttccaacag 2346
QY 2711 tgatctgtatcttagtttcaattatttcttcttgatgatatgacataataaaattt 2770
Db 2347 tgatctgtatcttagtttcaattatttcttcttgatgatatgacataataaaattt 2406
QY 2771 cattgagaatttctcaattgtatctagttaaatagcacagtttgaaaacttgtctgaga 2830
Db 2407 cattgagaatttctcaattgtatctagttaaatagcacagtttgaaaacttgtctgaga 2466
QY 2831 ctgactttatcaataatctaacccgacaagaatcatatccatgtgtatgtgtgaacatt 2890
Db 2467 ctgactttatcaataatctaacccgacaagaatcatatccatgtgtatgtgtgaacatt 2526
QY 2891 ttatttcaattgactaaacccagacagtttcagtgatgcaaatgtgtgcctctggttc 2950
Db 2527 ttatttcaattgactaaacccagacagtttcagtgatgcaaatgtgtgcctctggttc 2586
QY 2951 agctgaaacagtcctgtgacttcaaaaacctgtaataagtcctccacagttgtataaatt 3010
Db 2587 agctgaaacagtcctgtgacttcaaaaacctgtaataagtcctccacagttgtataaatt 2646
QY 3011 ggacaatttaggaattttaacttagatgatcatgttggttccatttttatattcatttt 3070

Db 2647 ggacaatttaggaattttaacttagatgatcatgttggttccatttttatattcatttt 2706
QY 3071 attttgttaaltgcaaacaggaacttaaatgaaacttgatctctgttttaagaattattaa 3130
Db 2707 attttgttaaltgcaaacaggaacttaaatgaaacttgatctctgttttaagaattattaa 2766
QY 3131 aaacatgtgtatctatacatatggtccttgaggaacttagctttcaactacacacagga 3190
Db 2767 aaacatgtgtatctatacatatggtccttgaggaacttagctttcaactacacacagga 2826
QY 3191 tatgatctccatgttagtccatataaaacctgcagagtgatttccagagtgctcgatactg 3250
Db 2827 tatgatctccatgttagtccatataaaacctgcagagtgatttccagagtgctcgatactg 2886
QY 3251 ttaattcatctccattagggctgaaaagaatgaacctacglttctgtatacagctgtgt 3310
Db 2887 ttaattcatctccattagggctgaaaagaatgaacctacglttctgtatacagctgtgt 2946
QY 3311 gcttttgatgtgtgttactgtacacagaagtggtgcaactgaggctcgtcggtgtgtcc 3370
Db 2947 gcttttgatgtgtgttactgtacacagaagtggtgcaactgaggctcgtcggtgtgtcc 3006
QY 3371 gtatgaaaacctgtgtagccctgcagtgtaagtactgtctccatcatgtgttaacgtgg 3430
Db 3007 gtatgaaaacctgtgtagccctgcagtgtaagtactgtctccatcatgtgttaacgtgg 3066
QY 3431 aattttctcccatgtgaatgtgataaaacttaagtgtttgtcatcataaattgttaatt 3490
Db 3067 aattttctcccatgtgaatgtgataaaacttaagtgtttgtcatcataaattgttaatt 3126
QY 3491 actaaaaaaaaaaaaa 3506
Db 3127 actaaaaaaaaaaaaa 3142

RESULT 4
Z93327
ID Z93327 standard; cDNA; 3260 BP.
XX
AC Z93327;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; ; human; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 409..1311
FT /*tag= a
FT /product= Stauflen protein
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
XX Moulard AJ;
XX WPI: 2000-246924/22.
XX P-PSDB: Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful

PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1b; 96pp; English.
XX
CC Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
SQ Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

Query Match 82.1%; Score 2879; DB 21; Length 3260;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2892; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 612 agaaagcataacccctactgtagaactaatgcactgtgcatgaaacttgaaaaaac 671
DB 369 agaaagcataacccctactgtagaactaatgcactgtgcatgaaacttgaaaaaac 428
QY 672 aatgtataagcctgttgacccttactctcggatgcagtcacacctataactacaatgag 731
DB 429 aatgtataagcctgttgacccttactctcggatgcagtcacacctataactacaatgag 485
QY 732 aggaagtgtatcccccgagtgactttaccacattccagttccaccttacttatca 791
DB 486 aggaagtgtatcccccgagtgactttaccacattccagttccaccttacttatca 545
QY 792 agtgaacttctgtgaggagagacagcaatttaatgcaagaagaaagacacagagctgc 851
DB 546 agtgaacttctgtgaggagagacagcaatttaatgcaagaagaaagacacagagctgc 605
QY 852 gaaacacgattgtctgcccacaaagcgttgagatcctgcagaaatgagccctgcagagag 911
DB 606 gaaacacgattgtctgcccacaaagcgttgagatcctgcagaaatgagccctgcagagag 665
QY 912 gctggaggtgaatggaagagaatccggaagaagaatctcaataatctgaataagtca 971
DB 666 gctggaggtgaatggaagagaatccggaagaagaatctcaataatctgaataagtca 725
QY 972 agtgtttgagatgcacttaaacggaactgcctgtgaatttcgagggtgcccggagag 1031
DB 726 agtgtttgagatgcacttaaacggaactgcctgtgaatttcgagggtgcccggagag 785
QY 1032 tggccacccacatagaagaacttgtgaccaaagtttcggttgaggagttgtggggga 1091
DB 786 tggccacccacatagaagaacttgtgaccaaagtttcggttgaggagttgtggggga 845
QY 1092 agtgtgaagggaagaagaagaatttcaagaaaaaatgcccgcataagctgttcttgaga 1151
DB 846 agtgtgaagggaagaagaagaatttcaagaaaaaatgcccgcataagctgttcttgaga 905
QY 1152 gctgaagaagtaccgcccctgcctgcagttgaaagtaagcctagaatcaaaaagaa 1211
DB 906 gctgaagaagtaccgcccctgcctgcagttgaaagtaagcctagaatcaaaaagaa 965
QY 1212 aacaaaaaccatagtcaagccacagacaagcccagaatatgcccaggggatcaatccgat 1271
DB 966 aacaaaaaccatagtcaagccacagacaagcccagaatatgcccaggggatcaatccgat 1025

QY 1272 taagcactgcccagatccagcagcacaagaagagaagagccagagtlacacgtcct 1331
DB 1026 taagcactgcccagatccagcagcacaagaagagaagagccagagtlacacgtcct 1085
QY 1332 cacagagcagagcctcccgccgcccagggagttgtgatgcaggtgaaggttgaaacca 1391
DB 1086 cacagagcagagcctcccgccgcccagggagttgtgatgcaggtgaaggttgaaacca 1145
QY 1392 cactgcagaagaaggggcaccacaagaaggtggccaagcgcattgcagccgagaacat 1451
DB 1146 cactgcagaagaaggggcaccacaagaaggtggccaagcgcattgcagccgagaacat 1205
QY 1452 gctgagatcctgtgttcaaaagttcccgccgagcagcccaaccgcactcaagtc 1511
DB 1206 gctgagatcctgtgttcaaaagttcccgccgagcagcccaaccgcactcaagtc 1265
QY 1512 agagagaagacaccataaagaacacagggatggaagaaagttaaccttttgaaacc 1571
DB 1266 agagagaagacaccataaagaacacagggatggaagaaagttaaccttttgaaacc 1325
QY 1572 tggctctgggatgaaaaatgaggactagtataaaggatgattccagatgccttatct 1631
DB 1326 tggctctgggatgaaaaatgaggactagtataaaggatgattccagatgccttatct 1385
QY 1632 aagtcatcagcagcctgcctgtcgtgaattcttcccatgtgtcccgagggtcgccagctgt 1691
DB 1386 aagtcatcagcagcctgcctgtcgtgaattcttcccatgtgtcccgagggtcgccagctgt 1445
QY 1692 aggaagtatgcaaggacatcacaccaagaatttaccagggcagctccgaatcctgcca 1751
DB 1446 aggaagtatgcaaggacatcacaccaagaatttaccagggcagctccgaatcctgcca 1505
QY 1752 ggcacaggtlaactgcatgatatagcccgagggttgttatgggggacactcgccacagc 1811
DB 1506 ggcacaggtlaactgcatgatatagcccgagggttgttatgggggacactcgccacagc 1565
QY 1812 cgagaccattttaagaataacatctcttcaaggccagctaccccatggaacctctcacgag 1871
DB 1566 cgagaccattttaagaataacatctcttcaaggccagctaccccatggaacctctcacgag 1625
QY 1872 accctctgagcaactggaactatcttccagaagttccaggatccaggttgaatacaaga 1931
DB 1626 accctctgagcaactggaactatcttccagaagttccaggatccaggttgaatacaaga 1685
QY 1932 ctccccaacaaacaaagaacgaattgtatctcttatacaattgtctcctcagccacc 1991
DB 1686 ctccccaacaaacaaagaacgaattgtatctcttatacaattgtctcctcagccacc 1745
QY 1992 tctgatcagccatggtatcggcaagatgtgagttcctgcgatgatatgctgcgtgaa 2051
DB 1746 tctgatcagccatggtatcggcaagatgtgagttcctgcgatgatatgctgcgtgaa 1805
QY 2052 catcttaagtgtcgtctgagtgtgacccaacaaagtacagagatgccaagaacagaa 2111
DB 1806 catcttaagtgtcgtctgagtgtgacccaacaaagtacagagatgccaagaacagaa 1865
QY 2112 cggaccaatgtctgtgtgtggaggtgtgaaccttttctgccaatgaacctataaa 2171
DB 1866 cggaccaatgtctgtgtgtggaggtgtgaaccttttctgccaatgaacctataaa 1925
QY 2172 tcccaacatatatactgaaataactgaaactgtcttgaaaatttgyaatttctgatacct 2231
DB 1926 tcccaacatatatactgaaataactgaaactgtcttgaaaatttgyaatttctgatacct 1985
QY 2232 ccagtgggccgagagacaacggttggtaagagtgtggcagcagcagggaagacaacaga 2291
DB 1986 ccagtgggccgagagacaacggttggtaagagtgtggcagcagcagggaagacaacaga 2045
QY 2292 aacacaagagggcgtgtgtggccgctggaactgtgtgggttgtgtatggtccactc 2351
DB 2046 aacacaagagggcgtgtgtggccgctggaactgtgtgggttgtgtatggtccactc 2105

OY 2352 ggtgacctggcggtccctaagcaatagcagctgcctgtggygaagaaggctgccagcc 2411
Db 2106 ggtgacctggcggtccctaagcaatagcagctgcctgtggygaagaaggctgccagcc 2165
OY 2412 agctggtctcccggygacacagcagatccacacctgggcaacctcgtgtgtgtcttt 2471
Db 2166 agctggtctcccggygacacagcagatccacacctgggcaacctcgtgtgtgtcttt 2225
OY 2472 ttttccctgtgtgaagaagaacacgacgaccccccttctcaagctgtgctcaactagac 2531
Db 2226 ttttccctgtgtgaagaagaacacgacgaccccccttctcaagctgtgctcaactagac 2285
OY 2532 acattgggacaaaccttgacacagccatgcccagagaggcctttgacccgccccagagct 2591
Db 2286 acattgggacaaaccttgacacagccatgcccagagaggcctttgacccgccccagagct 2345
OY 2592 aaaaagcaccagagaaaatacaaatgcttcctactcagcgltgacccaactttctagtgtgc 2651
Db 2346 aaaaagcaccagagaaaatacaaatgcttcctactcagcgltgacccaactttctagtgtgc 2405
OY 2652 caccgccccaccacctcctgagtlaccacacacatcaccaactgcttctctccaacagt 2711
Db 2406 caccgccccaccacctcctgagtlaccacacacatcaccaactgcttctctccaacagt 2465
OY 2712 gatcgtatctctagtctcaattatcttctttgatgatatagacataataaaatttc 2771
Db 2466 gatcgtatctctagtctcaattatcttctttgatgatatagacataataaaatttc 2525
OY 2772 atttgagaattctcgaattgtatctagttaaatagcacaglttgaaacttgtctgagac 2831
Db 2526 atttgagaattctcgaattgtatctagttaaatagcacaglttgaaacttgtctgagac 2585
OY 2832 tgactttatcaataatctaacccgacaaagatcatatccatgtgtatgtgttagacattt 2891
Db 2586 tgactttatcaataatctaacccgacaaagatcatatccatgtgtatgtgttagacattt 2645
OY 2892 ttatttcattgactaacccagagacaglttcagtgatgcaaatgtgtgccctctgttca 2951
Db 2646 ttatttcattgactaacccagagacaglttcagtgatgcaaatgtgtgccctctgttca 2705
OY 2952 gctgaaacagtcctggaacttcaaaaacctgaaataagtcctccacagltgtataaatg 3011
Db 2706 gctgaaacagtcctggaacttcaaaaacctgaaataagtcctccacagltgtataaatg 2765
OY 3012 gacaatttaggaattttaactttaagatgatcatcatgtgtccattttatttcatttta 3071
Db 2766 gacaatttaggaattttaactttaagatgatcatcatgtgtccattttatttcatttta 2825
OY 3072 ttttgttaatgcaaacagacttaaatgaactttgatctctgtttttaaagattataa 3131
Db 2826 ttttgttaatgcaaacagacttaaatgaactttgatctctgtttttaaagattataa 2885
OY 3132 aaacattgtlatctatacatatgctcttgaggacttagcttcaactacatacaggat 3191
Db 2886 aaacattgtlatctatacatatgctcttgaggacttagcttcaactacatacaggat 2945
OY 3192 atgattccatgtagtccataataaacctgcagagtgatttccagagtgctcgatactgt 3251
Db 2946 atgattccatgtagtccataataaacctgcagagtgatttccagagtgctcgatactgt 3005
OY 3252 taatcacatccattagggcgtgaaagaatgacctagcttctgtatacagctgtgttg 3311
Db 3006 taatcacatccattagggcgtgaaagaatgacctagcttctgtatacagctgtgttg 3065
OY 3312 ctttgatgttgttactgtacacacagaaagtgtgtgacactgaggtctgcgtgtgtccg 3371
Db 3066 ctttgatgttgttactgtacacacagaaagtgtgtgacactgaggtctgcgtgtgtccg 3125
OY 3372 tatgaaaaaacctgtagccctgcgaggttaagtactgtcttccattcatgtttaacgtgga 3431
Db 3126 tatgaaaaaacctgtagccctgcgaggttaagtactgtcttccattcatgtttaacgtgga 3185
OY 3432 attttctcccatgtaagttaagtaaaacttaagtgttgtcatcaataaatgylaata 3491

Db 3186 attttctcccatgtaagttaagtaaaacttaagtgttgtcatcaataaatgylaata 3245
OY 3492 ctaaaaaaaaaa 3506
Db 3246 ctaaaaaaaaaa 3260

RESULT 5
X90786
ID X90786 standard; DNA; 3190 BP.
XX
AC X90786;
XX
DT 13-JAN-2000 (first entry)
XX
DE DNA encoding human staufen protein.
KW Human staufen protein; hStau; protein lysate; lung; kidney; testis;
KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;
KW telomerase assembly; maturation; transport; regulation; interaction;
KW inhibitory agent; antibody; immunoassay; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 345..1835
FT FT /*tag= a
FT FT /product= "Human staufen protein"
FT FT /note= "Binds to telomerase RNA"
FT FT 3172..3177
FT FT /*tag= b
XX
PN W09951255-A1.
XX
PD 14-OCT-1999.
XX
PF 06-APR-1999; 99WO-US07533.
XX
PR 06-APR-1998; 98US-0080783.
XX
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Greider CW, Le S;
XX
DR WPI; 1999-620168/53.
DR P-PSDB; Y26335.
XX
PT Human staufen polypeptide useful in methods for identifying telomerase
PT inhibitors -
XX
PS Claim 13; Page 41-42; 50pp; English.
XX
CC The present sequence is an isolated DNA encoding human staufen (hStau)
CC protein. The hStau protein was detected in protein lysates from lung,
CC kidney, testis and ovary, but not in brain and heart. It binds
CC specifically to human telomerase RNA and can complex with the
CC telomerase catalytic subunit. The hStau protein may play a role in
CC telomerase assembly, maturation, transport and regulation. Interaction
CC between hStau and telomerase can be inhibited by administering inhibitory
CC agents to the cell. These telomerase inhibitors can be identified using
CC hStau protein. Antibodies generated against hStau can be used in various
CC immunoassays.
XX
SQ Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;

Query Match 81.7%; Score 2866; DB 20; Length 3190;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2880; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
OY 612 agaaagcataaccctactgttagaactaatgacactgtgcatgaaacttggaataaac 671
|||||

Db 305 agaaagcataccctactgttagaactaatgcactgtgcatgaaaacttggaaaaaacc 364
QY 672 aatgtataagcctgtgtgacccttactctcggatgcagtccacctaatacacatatgag 731
Db 365 aatgtataagcctgtgtgacccttactctcggatgcagtccacctaatacacatatgag 424
QY 732 aggaggtgcttatcccccggaggtacttttaccatttccagttccaccttacttatca 791
Db 425 aggaggtgcttatcccccggaggtacttttaccatttccagttccaccttacttatca 484
QY 792 agtggaaacttctgtgaggagagcagcgaatttaatgccaagaagaaagacagagctgc 851
Db 485 agtggaaacttctgtgaggagagcagcgaatttaatgccaagaagaaagacagagctgc 544
QY 852 gaaacacgatatgctgtgccaaaagcgttgagatccctgcagaatgagccctgcagagag 911
Db 545 gaaacacgatatgctgtgccaaaagcgttgagatccctgcagaatgagccctgcagagag 604
QY 912 gctggaagtgaatggaagagaatccgaaagaagaaatctcaataatctgaataagtca 971
Db 605 gctggaagtgaatggaagagaatccgaaagaagaaatctcaataatctgaataagtca 664
QY 972 agtgtttgagattgcacttaaacggaactgtgctgtgaatttcgaggtggtgcccggagag 1031
Db 665 agtgtttgagattgcacttaaacggaactgtgctgtgaatttcgaggtggtgcccggagag 724
QY 1032 tggcccaccccacatagaaagaactttgtgaccaaagtttcggttgaggagtttgtgggga 1091
Db 725 tggcccaccccacatagaaagaactttgtgaccaaagtttcggttgaggagtttgtgggga 784
QY 1092 aggtgaaaggaaagcaagaagaattccaagaaaaaatgcccacatagctgttcttgagga 1151
Db 785 aggtgaaaggaaagcaagaagaattccaagaaaaaatgcccacatagctgttcttgagga 844
QY 1152 gctgaaagaagttaccgccccctgctgtcagttgaaacgagtaaaagcctagaaatcaaaaaagaa 1211
Db 845 gctgaaagaagttaccgccccctgctgtcagttgaaacgagtaaaagcctagaaatcaaaaaagaa 904
QY 1212 aacaaaaaccatagtcgaagccacagacaaagcccagaataatgycacaggggatcaatccgat 1271
Db 905 aacaaaaaccatagtcgaagccacagacaaagcccagaataatgycacaggggatcaatccgat 964
QY 1272 tagccgacttggcccaagatccagcagggcaaaaaaggaaggaagccagagtacacgctcct 1331
Db 965 tagccgacttggcccaagatccagcagggcaaaaaaggaaggaagccagagtacacgctcct 1024
QY 1332 cacagaagcgaagcctcccgccgcccgaaggaagtttgtatgcaagtgaaaggttgaaacca 1391
Db 1025 cacagaagcgaagcctcccgccgcccgaaggaagtttgtatgcaagtgaaaggttgaaacca 1084
QY 1392 caactgcagaagaacgaggcaaccaagaaggtgyccaagcgcaatgcaagcgagaacat 1451
Db 1085 caactgcagaagaacgaggcaaccaagaaggtgyccaagcgcaatgcaagcgagaacat 1144
QY 1452 gctggaagatccttggtttcaaaagttcccgagcggcagcccacaaacccgcaactcaagtc 1511
Db 1145 gctggaagatccttggtttcaaaagttcccgagcggcagcccacaaacccgcaactcaagtc 1204
QY 1512 agaggaagaagacaccataaagaacccagggagtgaagaaagtaacctttttgaacc 1571
Db 1205 agaggaagaagacaccataaagaacccagggagtgaagaaagtaacctttttgaacc 1264
QY 1572 tggctctgaggatgtaaaaatgagactagtaataaagagatgagttcaggatgccttatct 1631
Db 1265 tggctctgaggatgtaaaaatgagactagtaataaagagatgagttcaggatgccttatct 1324
QY 1632 aagtcatacagcagctgcctgtctgaatcttcccatgtgtgcccaaggtcgccagagctgt 1691
Db 1325 aagtcatacagcagctgcctgtctgaatcttcccatgtgtgcccaaggtcgccagagctgt 1384
QY 1692 aggaagttagtcaaggaacatcacaccaagaattttaccagggcagctccgaatcctgcgcaa 1751
Db 1385 aggaagttagtcaaggaacatcacaccaagaattttaccagggcagctccgaatcctgcgcaa 1444

QY 1752 ggccacggtaaactgccaatgatagcccgagagttgttgtatggtgggcaacctgcgccacagc 1811
Db 1445 ggccacggtaaactgccaatgatagcccgagagttgttgtatggtgggcaacctgcgccacagc 1504
QY 1812 cgaagacattttaagaataacaatctctcgaaggccaagtaacccaatggaacctctcagag 1871
Db 1505 cgaagacattttaagaataacaatctctcgaaggccaagtaacccaatggaacctctcagag 1564
QY 1872 acccttgagcaactggaactaacttccagagttccaggaatccagglttgaatalacaaga 1931
Db 1565 acccttgagcaactggaactaacttccagagttccaggaatccagglttgaatalacaaga 1624
QY 1932 ctcccccaaaaaacaagaacgaatttgtatctcttacaattgtcctctcagccacc 1991
Db 1625 ctcccccaaaaaacaagaacgaatttgtatctcttacaattgtcctctcagccacc 1684
QY 1992 tctgatcagccaatggtatcggcaagatgtygagttcctgcacatgatatggtgcgtgaa 2051
Db 1685 tctgatcagccaatggtatcggcaagatgtygagttcctgcacatgatatggtgcgtgaa 1744
QY 2052 catcttaaaagtgtgtctgtgagttggaaccaaaagtlacagagatgccaagaacagga 2111
Db 1745 catcttaaaagtgtgtctgtgagttggaaccaaaagtlacagagatgccaagaacagga 1804
QY 2112 cggaccaatgtctgtgtgtgaggagtgctgaaaccttctgtgccaatgaaacattataaa 2171
Db 1805 cggaccaatgtctgtgtgtgaggagtgctgaaaccttctgtgccaatgaaacattataaa 1864
QY 2172 tcccaacatatatactgaaataactgaaacgtcttgaaaatlttgaaatttctgataacct 2231
Db 1865 tcccaacatatatactgaaataactgaaacgtcttgaaaatlttgaaatttctgataacct 1924
QY 2232 ccagtggtgcccgaaggaacacggtgtggttaaagatgttggcagcagcagggaagacaacaga 2291
Db 1925 ccagtggtgcccgaaggaacacacggtgtggttaaagatgttggcagcagcagggaagacaacaga 1984
QY 2292 aacacaaggaagcggctgtggcc-ggctggaactgtgtggtgttgtgtatggccact 2350
Db 1985 aacacaaggaagcggctgtggccggcggtggaactgtgtggtgttgtgtatggccact 2044
QY 2351 cggltgaacctgtgcccctaagcaatagcagctgcctgttggggaagaaaggtgcgccagc 2410
Db 2045 cggltgaacctgtgcccctaagcaatagcagctgcctgttggggaagaaaggtgcgccagc 2104
QY 2411 cagctgttctcccggaacacacagcagatccacacaccttgggcacctcgttgtgtctt 2470
Db 2105 cagctgttctcccggaacacacagcagatccacacaccttgggcacctcgttgtgtctt 2164
QY 2471 tttttccctgtgtgaaagaagaacggcagaccccttctcaagcttggtcactcaga 2530
Db 2165 tttttccctgtgtgaaagaagaacggcagaccccttctcaagcttggtcactcaga 2224
QY 2531 cacattgggacaacaccttgacagccatgcccagagagaagccttggaccgccccagagc 2590
Db 2225 cacattgggacaacaccttgacagccatgcccagagagaagccttggaccgccccagagc 2284
QY 2591 taaaagcaccaggaataatcaaatgcttctactcagcgtgacccaacttttctagtgtg 2650
Db 2285 taaaagcaccaggaataatcaaatgcttctactcagcgtgacccaacttttctagtgtg 2344
QY 2651 ccaagggccccaacacctcctgcagtlacccaacacatcaaccaactgttctctccaag 2710
Db 2345 ccaagggccccaacacctcctgcagtlacccaacacatcaaccaactgttctctccaag 2404
QY 2711 tgatctgtattcttagtttcatatttcttlttgatgtgatalgacactataaaatltt 2770
Db 2405 tgatctgtattcttagtttcatatttcttlttgatgtgatalgacactataaaatltt 2464
QY 2771 catttgaagaatttctcaattgttatctagttaaaatagcacagtttggaaactgtctgaga 2830
Db 2465 catttgaagaatttctcaattgttatctagttaaaatagcacagtttggaaactgtctgaga 2524

QY 2831 ctgaccttatcaataactaacgcgacaagaatcatatccatgtgtatgtgtagacatt 2890
|||||
Db 2525 ctgaccttatcaataactaacgcgacaagaatcatatccatgtgtatgtgtagacatt 2584
QY 2891 ttatttcattgactaacccagagacagtttcagtgatgcaaatgtgtgcccctgtgtc 2950
|||||
Db 2585 ttatttcattgactaacccagagacagtttcagtgatgcaaatgtgtgcccctgtgtc 2644
QY 2951 agctgaacacagtcctgtgactttcaaaaaccttgaaataagtcctccacagttgtataatt 3010
|||||
Db 2645 agctgaacacagtcctgtgactttcaaaaaccttgaaataagtcctccacagttgtataatt 2704
QY 3011 ggacaatttaggaattttaaactttagatgcatcatttggtccatttttattcattttt 3070
|||||
Db 2705 ggacaatttaggaattttaaactttagatgcatcatttggtccatttttattcattttt 2764
QY 3071 atttttgttaatgcaaaaggacttaaatgaactttgactctgtgttttaagaattataa 3130
|||||
Db 2765 atttttgttaatgcaaaaggacttaaatgaactttgactctgtgttttaagaattataa 2824
QY 3131 aaaacattgtgtatctatacatatgctcttgagagacttagcttcactacactacagga 3190
|||||
Db 2825 aaaacattgtgtatctatacatatgctcttgagagacttagcttcactacactacagga 2884
QY 3191 tatgatctccatgtagtcataataaacctgcagagtgatttccagagtgctcgatactg 3250
|||||
Db 2885 tatgatctccatgtagtcataataaacctgcagagtgatttccagagtgctcgatactg 2944
QY 3251 ttaattacatctccatttagggctgaaaagaatgacctacgcttctgtatacagctgtgtt 3310
|||||
Db 2945 ttaattacatctccatttagggctgaaaagaatgacctacgcttctgtatacagctgtgtt 3004
QY 3311 gcttttgatgtgtgttactgtacacagagaagtgtgtgcactgaggtcgtgtgtgtcc 3370
|||||
Db 3005 gcttttgatgtgtgttactgtacacagagaagtgtgtgcactgaggtcgtgtgtgtcc 3064
QY 3371 gtatggaaaaacctggtagccctgcgagttaaagtactgtctccatttgatttaacgtg 3430
|||||
Db 3065 gtatggaaaaacctggtagccctgcgagttaaagtactgtctccatttgatttaacgtg 3124
QY 3431 aattttctcccatgtgaagttaaaaaacttaagtgttgtgcatcaataaatgttaat 3490
|||||
Db 3125 aattttctcccatgtgaagttaaaaaacttaagtgttgtgcatcaataaatgttaat 3184
QY 3491 actaaa 3496
|||
Db 3185 acttaa 3190

RESULT 6
293328
ID 293328 standard; cDNA; 2859 BP.
XX
AC 293328;

XX 04-JUL-2000 (first entry)
XX
DE Mouse staufen cDNA.
XX

KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; mouse; ss.
XX

OS Mus musculus.
XX

XX Key Location/Qualifiers
FH CDS 325..1788
FT CDS /*tag= a
FT /product= Staufen protein
XX

PN CA2238656-A1.
XX
PD 22-NOV-1999.

22-NOV-1999.

XX 22-MAY-1998; 98CA-2238656.
PF
XX 22-MAY-1998; 98CA-2238656.
PR
XX (UYMO-) UNIV MONTREAL.
PA
XX Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;
PI Moulard AJ;
PI
XX WPI; 2000-246924/22.
DR P-PSDB; Y83024.
DR
XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1c; 96pp; English.
PS
XX Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
SQ Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;

Query Match 35.8%; Score 1256; DB 21; Length 2859;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1420; Conservative 0; Mismatches 210; Indels 8; Gaps 3;

QY 611 cagaagcataaccctactgtagaactaaatgacatgcatgaaacttgaaaaaac 670
|||||
Db 263 cagaagcataagcttctaccatgagctcaatgacatgctgtgtgaacttgaaaaaac 322
QY 671 caatgtataagcctgttgacccttactctcgatgcagtcacactataactacaatatga 730
|||||
Db 323 caatgtataagcctgttgacccttactctcgatgcagtcacactacaactatgycatgc 382
QY 731 gagaggtgctatcccccgagtaactttaccattccagttccaccttactatc 790
|||||
Db 383 gtgaggtgctatcccccgagtaactttaccattccagttccaccttactatc 442
QY 791 aagtgaacttctgtgagagcagcaacttaatgcaaggaagacaagacagctg 850
|||||
Db 443 aagtgaacttctgtgagagcagcaacttaatgcaaggaagacaagacagctg 502
QY 851 cgaacacagatgctgtgccaaagcgttgagatcctgcagaatgagccctgcagaga 910
|||||
Db 503 tgaacacagatgcccctgtccgtgcgtgaggaactctgcagagtgaacccctgcagaaa 562
QY 911 ggctgagagtgaatggaagagatccgaagaagaaatctcaataatctgaataagtc 970
|||||
Db 563 ggttgagagtgaatggaagagatccgaagaagaaatctcaataatctgaataagtc 622
QY 971 aagtgttgagattgcaacttaacggaacttgctgtgaatttcgaggtgcccgggaga 1030
|||||
Db 623 aagtgttgaaattgcgtgagcggaaatttgctgtgaatttgaggtgcccgggaga 682
QY 1031 gtgcccacccacatgaagaacttgtgaccaaggttcggttgaggagttgtgggg 1090
|||||
Db 683 gtgcccacccacatgaagaacttgtgaccagaggtttcagttgaggagttgtgagg 742

QY 809 gagacagcaatttaatgycaaagaaagacacagcagctgcgaacacgatgtctg 868
|||||
Db 1 gagacagcaatttaatgycaaagaaagacacagcagctgcgaacacgatgtctg 60
QY 869 ccaaaagcgttgagatccctgcagaaatgagccctgcagagagagctggaagtgaatgaa 928
|||||
Db 61 ccaaaagcgttgagatccctgcagaaatgagccctgcagagagagctggaagtgaatgaa 120
QY 929 gagaatccgaagaagaataatcaataatctgaataaagtcaagtgtttgagattgcac 988
|||||
Db 121 gagaatccgaagaagaataatcaataatctgaataaagtcaagtgtttgagattgcac 180
QY 989 ttaaacggaacttgctgtgaatttcgaggtgcccgggagagtggccacccacatga 1048
|||||
Db 181 ttaaacggaacttgctgtgaatttcgaggtgcccgggagagtggccacccacatga 240
QY 1049 agaacttgtgaccaagtttcggttggtgggagttgtggggaaagtgaagggaagaagca 1108
|||||
Db 241 agaacttgtgaccaagtttcggttggtgggagttgtggggaaagtgaagggaagaagca 300
QY 1109 agaagatttcaagaaataatgccgcataagctgttcttgagagagctgaagaagttaaccgc 1168
|||||
Db 301 agaagatttcaagaaataatgccgcataagctgttcttgagagagctgaagaagttaaccgc 360
QY 1169 ccctgcctgcagttgacgaagtaaaagcctagaatcaaaaaagaaaaaacccatagtca 1228
|||||
Db 361 ccctgcctgcagttgacgaagtaaaagcctagaatcaaaaaagaaaaaacccatagtca 420
QY 1229 agccacagacaagcccagaatatgcccaggggatccaatccgattagccgactggcccaga 1288
|||||
Db 421 agccacagacaagcccagaatatgcccaggggatccaatccgattagccgactggcccaga 480
QY 1289 tccagcagycaaaaaaggagaagagccagagtaacacgctcctcacagagcagggcctcc 1348
|||||
Db 481 tccagcagycaaaaaaggagaagagccagagtaacacgctcctcacagagcagggcctcc 540
QY 1349 cgcgccgcagggagtttgtgatgcaggtggaagttggaaccacactgcagaaggaacgg 1408
|||||
Db 541 cgcgccgcagggagtttgtgatgcaggtggaagttggaaccacactgcagaaggaacgg 600
QY 1409 gcaccaacaagaagtgcccaagcgcaatgcagccgagacaatgctggaatccttggt 1468
|||||
Db 601 gcaccaacaagaagtgcccaagcgcaatgcagccgagacaatgctggaatccttggt 660
QY 1469 tcaaaagtcgcgcagcgccaccaaacccgcactcaagtcagagggagaagacacca 1528
|||||
Db 661 tcaaaagtcgcgcagcgccaccaaacccgcactcaagtcagagggagaagacacca 720
QY 1529 taaagaaaccaggggatggaagaaagtaaccctttttgaacctggtcttgggatgaa 1588
|||||
Db 721 taaagaaaccaggggatggaagaaagtaaccctttttgaacctggtcttgggatgaa 780
QY 1589 atgggactagtataaagagatgagttcagagatgccttatctaagtcatcagcagctgc 1648
|||||
Db 781 atgggactagtataaagagatgagttcagagatgccttatctaagtcatcagcagctgc 840
QY 1649 ctgctggaattcttcccatggtgcccgagtgcccaggtctgtagagtagtcaaggaac 1708
|||||
Db 841 ctgctggaattcttcccatggtgcccgagtgcccaggtctgtagagtagtcaaggaac 900
QY 1709 atcacaccaagaattttacacgggcagctccgaatccctgcacaagycacgtaactgcca 1768
|||||
Db 901 atcacaccaagaattttacacgggcagctccgaatccctgcacaagycacgtaactgcca 960
QY 1769 tgatagcccgagagttgttgtatggtgggcacctgcgccacag-ccgagaccattttaag 1827
|||||
Db 961 tgatagcccgagagttgttgtatggtgggcacctgcgccacagcccgagaccattttaag 1020
QY 1828 aatacatctcttcaggcacagtaaccatgacctctcaacgagaccctctgaagcaactg 1887
|||||
Db 1021 aatacatctcttcaggcacagtaaccatgacctctcaacgagaccctctgaagcaactg 1080

QY 1888 gactatcttcagaggtccagggattccaggttgaataacaagact 1934
|||||
Db 1081 gactatcttcnagagtcagggattncaggttgaataacaagact 1127
RESULT 8
X40002
ID X40002 standard; DNA; 946 BP.
XX
AC X40002;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629-630; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other;

Query Match 22.8%; Score 798.8; DB 20; Length 946;
Best Local Similarity 93.6%; Pred. No. 7.4e-195;
Matches 888; Conservative 0; Mismatches 54; Indels 7; Gaps 6;

QY 809 gagacagcaatttaatgycaaagaaagacacagcagctgcgaacacgatgtctg 868
|||||
Db 1 gagacagcaatttaatgycaaagaaagacacagcagctgcgaacacgatgtctg 60

QY 869 ccaagcgttgagatcctgcagaatgagccctgcagagagcgtgaggtgaa 928
|||||
Db 61 ccaagcgttgagatcctgcagaatgagccctgcagagagcgtgaggtgaa 120
QY 929 gagaatccgaagaataatctcaataatctgaataagtcaagtgttgagattgcac 988
|||||
Db 121 gagaatccgaagaataatctcaataatctgaataagtcaagtgttgagattgcac 180
QY 989 ttaacggaacttgctgtgaatttcgaggtgcccggagagtgcccacccacatga 1048
|||||
Db 181 ttaacggaacttgctgtgaatttcgaggtgcccggagagtgcccacccacatga 240
QY 1049 agaacttgtagaccaaggttcggttgaggagtttggtgggaagtgaaaggaagca 1108
|||||
Db 241 agaacttgtagaccaaggttcggttgaggagtttggtgggaagtgaaaggaagca 300
QY 1109 agaagattcaagaataatgcgcgcatagcgtgtcttgagagcgtgaagaagtlaccgc 1168
|||||
Db 301 agaagattcaagaataatgcgcgcatagcgtgtcttgagagcgtgaagaagtlaccgc 360
QY 1169 ccctgcctgcagtgaacgagtaaacgctgaatcaaaaagaataaaccatagtca 1228
|||||
Db 361 ccctgcctgcagtgaacgagtaaacgctgaatcaaaaagaataaaccatagtca 420
QY 1229 agccacagacaagccagaaatatgcccaggggatcaatccgattagccgactgcccaga 1288
|||||
Db 421 agccacagacaagccagaaatatgcccaggggatcaatccgattagccgactgcccaga 480
QY 1289 tccagcagcgaagaaaggagaagagccagagtaacgcctcctcacagaagcaggcctcc 1348
|||||
Db 481 tccagcagcgaagaaaggagaagagccagagtaacgcctcctcacagaagcaggcctnc 540
QY 1349 cgcgcgcgcagggagttgtgatgcaggtgaaggttgaaaccacac-tgcagaaggaagc 1407
|||||
Db 541 cgcgcgcgcagggagttgtgatgcaggtgaaggttgaaaccacac-tgcagaaggaagc 600
QY 1408 ggcacacaagaaggtgccaagcgcgaatgcagccgagacaatgctggagatccttgt 1467
|||||
Db 601 ggcacacaagaaggtgccaagcgcgaatgcagccgagacaatgctggagatccttgt 660
QY 1468 ttc-aaagtcgcgcgcgcagccgcacccaacccgcactcaagtcagagaggaagacacc 1526
|||||
Db 661 ttc-aaagtcgcgcgcgcagccgcacccaacccgcactcaagtcagagaggaagacacc 720
QY 1527 cataaagaacccagggatggaagaaagtcaaccttttgaaacctggtcctgggagtg 1586
|||||
Db 721 attaaggaacccagggatggaagaaagtcaaccttttgaaacctggtcctgggagtg 778
QY 1587 aaatgaggactagtaataaagagatgagtcagatgacctatctaagtcacagcagct 1646
|||||
Db 779 aaatgagg-ctgtantaaagagatgagtcagatgagtcagatgagtcacagcagct 836
QY 1647 gcctgctggaattctcccatggtgcccgaaggtgcgcccaggtcgaaggttagtcaagc 1706
|||||
Db 837 tgcgtgctggaatctctcccatggtgcccgaaggtgcgcccaggtcgaaggttagtcaagc 895
QY 1707 acatcacaccaagaatttacagggcagctccgaatcctgcgaagggc 1755
|||||
Db 896 ccnttcncccaagaatttcccggcagtttccaatctgcgaagggc 944

RESULT 9
X40000
ID X40000 standard; DNA; 844 BP.
XX
AC X40000;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Oyata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

Query Match 21.6%; Score 755.6; DB 20; Length 844;
Best Local Similarity 96.3%; Pred. No. 8.1e-184;
Matches 812; Conservative 0; Mismatches 26; Indels 5; Gaps 4;

QY 971 aagtgcttgagatgcaacttaaacggaacttgctgtgaatttcgaaggtgcccggagga 1030
|||||
Db 1 aagtgcttgagatgcaacttaaacggaacttgctgtgaatttcgaaggtgcccggagga 60
QY 1031 gtggccaccccatgaagaacttgtagccaaggttcggtgggaggttggtgggg 1090
|||||
Db 61 gtggccaccccatgaagaacttgtagccaaggttcggtgggaggttggtgggg 120
QY 1091 aagtggaagggaagaagaagatttcaagaanaaatgcccagatagctgttcttgagg 1150
|||||
Db 121 aagtggaagggaagaagaagatttcaagaanaaatgcccagatagctgttcttgagg 180
QY 1151 agctgaagaagttaccgcccctgctgcagttgaaagagtaagcctagaatcaaaaaa 1210
|||||
Db 181 agctgaagaagttaccgcccctgctgcagttgaaagagtaagcctagaatcaaaaaa 240
QY 1211 aaacaaaccatagtcaagccacagacaagccagaaatatgcccaggggatcaatcga 1270
|||||
Db 241 aaacaaaccatagtcaagccacagacaagccagaaatatgcccaggggatcaatcga 300

QY 1271 ttagccgactggccagatccagcagcaaaaaaggagaagcagagtlacagctcc 1330
|||||
Db 301 ttagccgactggccagatccagcagcaaaaaaggagaagcagagtlacagctcc 360
QY 1331 tcacagagcagagcctcccgccgagcagaggtttgtatgcaggtgaagttggaacc 1390
|||||
Db 361 tcacagagcagagcctcccgccgagcagaggtttgtatgcaggtgaagttggaacc 420
QY 1391 acactgcagagaagcagcggccaccaacaagaaggtgcccagcgaatgcagccgagaaca 1450
|||||
Db 421 acactgcagagaagcagcggccaccaacaagaaggtgcccagcgaatgcagccgagaaca 480
QY 1451 tgcctgagatccttggtttcaaaagtcgcgagcgcagccaccaaccgcaactcaagt 1510
|||||
Db 481 tgcctgagatccttggtttcaaaagtcgcgagcgcagccaccaaccgcaactcaagt 540
QY 1511 cagagagagaagacacccataaagaacaggggatggaagaaaagtaacctttttgaac 1570
|||||
Db 541 cagagagagaagacacccataaagaacaggggatggaagaaaagtaacctttttgaac 600
QY 1571 ctgctctgggatgtaaaatgagactagtaataaagagatgagttcagagatgcctatc 1630
|||||
Db 601 ctgctctgggatgtaaaatgagactagtaataaagagatgagttcagagatgcctatc 660
QY 1631 taagtcatcagcagctgctgtctggaattcttcccatggtgccgaggtcgccagcgtg 1690
|||||
Db 661 taagtcatcagcagctgctgtctggaattcttcccatggtgccgaggtcgccagcgtg 720
QY 1691 tagg-aqtlagtcagagacatcacaccaaaagatttaccagggcagctccgaatcctgcc 1749
|||||
Db 721 taggaagttagtcagagacatcacaccaaaagatttaccagggcagctccgaatcctgcc 779
QY 1750 aagggcagcgttaactgcatgtagcccgagaggtgtgtatggtgggcacctgcgccaca 1809
|||||
Db 780 aagggcagcgttaactgcatgtagcccgagaggtgtgtatggtgggcacctgcgccaca 836
QY 1810 gcc 1812
| |
Db 837 ggc 839
RESULT 10
X39992
ID X39992 standard; DNA; 773 BP.
XX
AC X39992;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

Query Match 21.5%; Score 753.2; DB 20; Length 773;
Best Local Similarity 99.1%; Pred. No. 3.2e-183;
Matches 766; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1163 taccgcccctgctgcagttgacagagtaaaagcctagaaatcaaaaaagaacaaccca 1222
| | | | |
Db 1 taccgcccctgctgcagttgacagagtaaaagcctagaaatcaaaaaagaacaaccca 60
QY 1223 tagtcaagccacagacaagcccaaatatgcccaggggatccaatccgattagccgactgg 1282
| | | | |
Db 61 tagtcaagccacagacaagcccaaatatgcccaggggatccaatccgattagccgactgg 120
QY 1283 cccagatccagcagcgaacaaagagaagagcagcagatracacgctcctcacagagcgag 1342
| | | | |
Db 121 cccagatccagcagcgaacaaagagaagagcagcagatracacgctcctcacagagcgag 180
QY 1343 gcctcccgccgagcagggagtttgtatgcaggtgtaaggttggaaccacactgcagaag 1402
| | | | |
Db 181 gcctcccgccgagcagggagtttgtatgcaggtgtaaggttggaaccacactgcagaag 240
QY 1403 gaacggcaccacaagaaggtgcccgaagcgcaatgcagccgagaacatgctgagatcc 1462
| | | | |
Db 241 gaacggcaccacaagaaggtgcccgaagcgcaatgcagccgagaacatgctgagatcc 300
QY 1463 ttggttcaaaagtcgcgagcgcagccaccaaccgcaactcaagtcagagagaaga 1522
| | | | |
Db 301 ttggttcaaaagtcgcgagcgcagccaccaaccgcaactcaagtcagagagaaga 360
QY 1523 caccataaagaacacaggggatggaagaaagtaacctttttgaacctgctcgggg 1582
| | | | |
Db 361 caccataaagaacacaggggatggaagaaagtaacctttttgaacctgctcgggg 420
QY 1583 atgaaatgggactagtaataaagagatgagttcagatgcttatctaagtcacagc 1642
| | | | |
Db 421 atgaaatgggactagtaataaagagatgagttcagatgcttatctaagtcacagc 480
QY 1643 agctgcctgctggaattcttcccatggtgccgaggtcgcccaggtctgtagagttagtc 1702
| | | | |
Db 481 agctgcctgctggaattcttcccatggtgccgaggtcgcccaggtctgtagagttagtc 540
QY 1703 aagacatcacaccaaagattttaccagggcagctccgaatcctgccaagggcaggttaa 1762
| | | | |
Db 541 aagacatcacaccaaagattttaccagggcagctccgaatcctgccaagggcaggttaa 600

QY 1763 ctgccatgatagccgagagtgtgtatgaggcgaccctgcgccacag-ccgagaccatt 1821
|||||
Db 601 ctgccatgatagcccgagagtgtgtatgaggcgaccctgcgccacagcccgagaccatt 660
QY 1822 ttaagaataacatctcttcaggccacgtaccatgtgacctctcacgagagaccctctgag 1881
|||||
Db 661 ttaagaataacatctcttcaggccacgtaccatgtgacctctcacgagagaccctntgag 720
QY 1882 caactgactatcttccagagagtcaggatccagggttgataatacaagactt 1934
|||||
Db 721 caactgactatcttccagagagtcaggatccagggttgataatacaagactt 773

RESULT 11
X40001/c
ID X40001 standard; DNA; 825 BP.
XX
AC X40001;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer.associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 825 BP; 289 A; 152 C; 151 G; 228 T; 5 other;

Query Match 15.7%; Score 551; DB 20; Length 825;
Best Local Similarity 95.9%; Pred. No. 2.1e-131;
Matches 585; Conservative 0; Mismatches 23; Indels 2; Gaps 2;
QY 2889 ttttattcatgtactaacccagagacagtttcagtgatgcaaatgtgtgccctcgtg 2948
||| ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 791 TTTAATTTCTTGCTTAACCCAGGNCAGTTCAGTGATGCAAAATGGGGCCNTCTGGT 732
QY 2949 tcagctgaaa-cagtcctgacttccaanaacctgtaataagtcctccacagtgatataa 3007
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 731 TCAGCTGAACACAGCTCTGGAAGCTTCAAAAACCTTGAATAAAGTCTCCCCAGTTGTATAA 672
QY 3008 attgacaatttaggaatttaaactttagatgatcattgtgtccatttattcatt 3067
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 ATTGAGACCATTTAGGAATTTTAACCTTAGATGATCATTTGGTCCATTTTATTNTCATTT 612
QY 3068 ttatttltgttaatgcaaacacagagac-ttaaatgaacttgcattctcgttlttaagatta 3126
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 TTTATTTTGTGTTAATGCAACAGGACTTTAAATGAACCTTGATCTCTGTTTAAAGATTA 552
QY 3127 ttaaaaaacattgtatctatatacatatgctcttgaggactttagcttccactac 3186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 TTAATAAACATTGTGTATCTATACATATAGGCTCTGAGGACTTAGCTTCACTACACTAC 492
QY 3187 aggatatgatctccatgtagtcataataaacctgcagagtgatlttccagagtgctcgat 3246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 491 AGGATATGATCTCCATGTAGTCCATATAAAGCTGCAGAGTGATTTTCCAGAGTGCTCGAT 432
QY 3247 actgttaattacatctccatttaggctgaaagaatgacctaagttctgtatacagctg 3306
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 431 ACTGTAATTACATCTCCATTAGGGCTGAAGAAGATGACCTACGTTCTGTATACAGCTG 372
QY 3307 tgttgcctttagatgtgtgttactgtacacagaagtgtgtgcactgaggtctcgtgtg 3366
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 TGTGCTTTGATGTGTGTACTGTACACAGAGTGTTGACACTGAGGCTCTGCGTGTG 312
QY 3367 gtccgtatggaanaacctgtagccctgcgagttgaagtactgtcttcattcatgtttacg 3426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 GTCCGTATGGAAGCCTGTGAGCCCTGCGAGTTGAAGTACTGCTCCATTCATTGTTTACG 252
QY 3427 ctggaatttctcctccatggaatgtaagtaaaaacttaagtgttgcatacaataatg 3486
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 CTGGAATTTTCTCCCATGGAATGTAAGTAATAACTTAAGTGTTCATCAATAAATG 192
QY 3487 taataactaa 3496
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 TAACTACTAAA 182

RESULT 12
X39993/c
ID X39993 standard; DNA; 794 BP.
XX
AC X39993;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
PF 15-JUL-1998; 98WO-US14679.
XX

PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Oyata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 794 BP; 273 A; 141 C; 152 G; 206 T; 22 other;

Query Match 14.3%; Score 501.4; DB 20; Length 794;
Best Local Similarity 96.3%; Pred. No. 1e-118;
Matches 517; Conservative 0; Mismatches 18; Indels 2; Gaps 1;
QY 2962 tcctcgacttcaaaaacctgaataagtcctccc--acagttgtataaattgacaattt 3019
DB 718 TCCTGACTTCAAAACCCTGAATAAGTCTCCCNCCAGTTGNATFAAATGGCCNATT 659
QY 3020 aggaatttaaaacttagatgatcattgtgtccattttatttattttttgtt 3079
DB 658 AGGAATTAAACNTTAGATGTTCATTGGTCCATTTTATTTCATTTTATTTTGGT 599
QY 3080 aatgcaaacaggaacttaaatgaacttgcctctgttttaagaattataaaacattg 3139
DB 598 AATGCAACACAGNCTTAATGAAGACTTGATCTCTGTTTAAAGATTATTAACACATTG 539
QY 3140 tgtatcatatcatatgtgctcttgaggacttagctttcaactacactacagatatgctc 3199
DB 538 TGNATCTATACATATGGCTCTTGAGGACTTAGCTTTCACACTACACTACAGATATGATCTC 479
QY 3200 catgtagtcacataaacaacctgcagagtgatttccagaagtgcgcgatactgtaattaca 3259
DB 478 CATGTAGTCCATATATAACCTGCAGAGTGATTTTCCANAGTGCTCGATACTGTTAATTACA 419
QY 3260 tcctcattaggcgtgaagaagaactacactacgtttctgtatatacagctgtgtctttgat 3319
DB 418 TCCTCCNTTAGGGCTGAAAAGATGACCTACGTTTCTGTATACANCTGTGTGCTTTGAT 359
QY 3320 gtgtgtactgtacacagaagtgtgtgactgaggtcctgcgtgtgtgtcgtatgga 3379
DB 358 GTTGTGTTACTGNACACAGAAGTGTGTGNCNTGAGGCTCTGCGTGTGTCGTAAGAAA 299
QY 3380 acctgttagccctgcgagttaagtactgtcttccattcaattgtttaagctggaattttct 3439
|||||

DB 298 GCCTGGTAGCCCTGCGAGTTAAGTACTGCTTCCATTGATTGTTACGCTGGAATTTTCT 239
QY 3440 ccccatggaatgtaagtaaaactaagtgttgcatacaataaatgtaactaaa 3496
DB 238 CCCCATGGAATGTAAGTAAACCTTAAGTGTGTCATCATNAATGGTAATACTAAA 182

RESULT 13
Z79981/C
ID Z79981 standard; cDNA; 660 BP.
XX
AC Z79981;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catlino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -
PS Claim 1; Page 158; 469pp; English.
XX
CC Z79917 to Z80766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other;

Query Match 12.4%; Score 433.8; DB 21; Length 660;
Best Local Similarity 92.4%; Pred. No. 1.9e-101;
Matches 495; Conservative 0; Mismatches 36; Indels 5; Gaps 4;
QY 1315 ccagagtacacgctcctcacagagcagcctccgcgcgcagggagttgtgtgcag 1374
DB 531 CCAAGGTACACGTTCTTCANGAAGCNAAGCCTTCGGNGCGNAGGAAGTTGTNATNCAG 472
QY 1375 gtgaagttggaacaacacactgcagaaggaacgycgacccaagaaggtggccaagcgc 1434
DB 471 GTNAAGGTTAGAACCCCATTTGCANNAAGGA--CGGGCACCAACAAGAAGGTGGCCAG-GC 415
QY 1435 aatgcagcgcgagaacatgctgtgagatccttgtttcaagtcgccgcagcgcgcacc 1494
|||||

Db 414 AATGCAG-CGAGANCATGTTGAAGATCCTGGGTTTCAAAAGTCCCGACGCGCAG-CCACC 357
QY 1495 aaaccgcactcaagtcagagagagaagacacccataaagaaccagggatggaagaaa 1554
Db 356 AAACCGGCACTCAAGTCAGAGAGAGAGACACCCATAAGAACCAGGGATGGAAGAAA 297
QY 1555 gtaacctttttgaacctggtctctggygatgaaaatggygactagtataaagagatgag 1614
Db 296 GTAACTTTTGAACCTGGCTCTGGGGATGAATAATGGGACTAGTATTAAGAGGATGAG 237
QY 1615 ttcaagatgacctatctaagtcatacagcagctgctctggaattctcccatggtgcc 1674
Db 236 TTCAGGATGCTTATCTAAGTCATCAGCAGCTGCTGCTGAATTCTTCCATGGTGCCC 177
QY 1675 gaagtcgcccagctgtagagagtagtcaagagacatcacaccaagaatttaccaggca 1734
Db 176 GAGTCGCCCGCAGGCTGTAGCAGTTAGTCAAGGACATCACACCAAGATTTTACGAGGCA 117
QY 1735 gctccgaatcctgcgaaggccacgtaactgcatgatatagcccgagagtgtgtatgag 1794
Db 116 GCTCCGAATCCTGCCAAGGCCACGGTAACCTGCATGATAGCCCGAGAGTTGTGTATGGG 57
QY 1795 ggcacctgcgccacagccgagaccattttaagaataacatctctcaggccacgt 1850
Db 56 GGCACCTCGCCACAGCCGAGACCATTTTAAAGAATAACATCTCTTCAGGCCACGT 1

RESULT 14

280269

ID 280269 standard; cDNA; 727 BP.

XX 280269;

DT 07-APR-2000 (first entry)

DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.

XX Human; gene expression product; diagnosis; tumour; colon cancer;

KW colorectal adenocarcinoma; cell line SW480; cell proliferation;

KW cytosstatic; sarcoma; breast cancer; neoplasia; dysplasia;

KW hyperplasia; ds.

KW

OS

XX Homo sapiens.

XX

PN WO9964576-A2.

XX 16-DEC-1999.

PD

PF 09-JUN-1999; 99WO-IB01062.

XX

PR 10-JUN-1998; 98US-0088801.

XX

PA (FARB) BAYER CORP.

XX

PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;

PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;

PI Schlegel R;

XX

XX

DR WPI; 2000-087220/07.

XX

PT Novel nucleic acids, used to develop products for the diagnosis and

PT treatment of disorders involving unwanted cell proliferation,

PT particularly cancers, especially colon cancer

XX

XX

PS Claim 15; Page 273; 469pp; English.

XX

CC 279917 to 280766 represent double stranded cDNA clones isolated from the

CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The

CC cDNA clones can be used to generate antisense oligonucleotides which

CC can be used for antisense therapy. Methods and products from the present

CC invention can be used for identifying and/or classifying cancerous cells

CC present in a human tumour, particularly in solid tumours, e.g. carcinomas

CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;

Query Match 11.8%; Score 412.2; DB 21; Length 727;
Best Local Similarity 95.5%; Pred. No. 7e-96;
Matches 444; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 752 ggtacttttaccattccagttccaccttaactttatcaagtgaactttctgtggag 811

Db 1 ggtacttttaccattccagttccaccttaactttatcaagtgaactttctgtggag 60

QY 812 gacagcaatttaatgcaaaaggaaagacagagctgcgaaacacagatgctgtgcc 871

Db 61 gacagcaatttaatgcaaaaggaaagacagagctgcgaaacacagatgctgtgcc 120

QY 872 aagcgtltgagatcctgcagaatgagccctgcagagagagctggaagtgaatggaag 931

Db 121 aagcgtltgagatcctgcagaatgagccctgcagagagagctggaagtgaatggaag 180

QY 932 aatccgaagaagaatctcaataaactgnaataaagtcagtggttgagattgcactta 991

Db 181 aatccgaagaagaatctcaataaactgnaataaagtcagtggttgagattgcactta 240

QY 992 aacggaacttgctgtgaatttcgagtggtgcccgggagagtggtgccacccacatga 1051

Db 241 aacggaacttgctgtgaatttcgagtggtgcccgggagagtggtgccacccacatga 300

QY 1052 actttgtaccaaggttcggttggtggagttgtggtgggaagtgaagggaagaaga 1111

Db 301 actttgtaccaaggttcggttggtggagttgtggtgggaagtgaagggaagaaga 360

QY 1112 agatttcaagaaaaatg-cgcgcataagctgttcttgaggagctgaagaagtlaccgcc 1170

Db 361 agatttcaagaaaaatgcccgcataagctgttcttgaggagctgaagaagtlaccgcc 420

QY 1171 ctgcctgcagttgaac---gagtaagcctagaatcaaaaagaaa 1212

Db 421 ctgcctgcagttgaac---gagtaagcctagaatcaaaaagaaa 465

RESULT 15

217411

ID 217411 standard; cDNA; 769 BP.

XX 217411;

AC

XX 12-OCT-1999 (first entry)

XX

DE Human gene expression product cDNA sequence SEQ ID NO:4884.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX

OS Homo sapiens.

XX

PN WO9938972-A2.

XX 05-AUG-1999.

PD

XX 28-JAN-1999; 99WO-US01619.

PF

XX 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX

PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crivenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
XX
PS Claim 1; Page 2322; 2479pp; English.

CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in Z12532 to Z17779. Also described is a
CC method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in Z12532 to Z17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.

XX
XX
SQ Sequence 769 BP; 176 A; 214 C; 135 G; 188 T; 56 other;

Query Match 10.5%; Score 369; DB 20; Length 769;
Best Local Similarity 82.3%; Pred. No. 8.5e-85;
Matches 502; Conservative 0; Mismatches 30; Indels 78; Gaps 4;

QY 121 ctccacggccactcgcctcttcctccctcgtccctctcctcctccttctcctt 180
Db 64 cacgagggccactcgcctcttcctccctccttcttctcctcctccttctcct 123
QY 181 ctccctccctcctcgcgcgacccgacggccggcggggagcgagctcgagc 240
Db 124 ctccctccctcctcgcgcgacccgacggccggcggggagcgagctcgagc 183
QY 241 agcagccagagttatataccacttaacctctcagaactgaacaagaacattgtcc 300
Db 184 agcagcc----- 190
QY 301 tggaaagccctctttaaagtagaacttagaactcatagcaactgaattaacctg 360
Db 191 -----agtagaacttagaactcatagcaactgaattaacctg 228
QY 361 cactgaaagctgtttactgcatlgttcaactlgttgaagtgacacgtctcaagt 420
Db 229 cactgaaagctgtttactgcatlgttcaactlgttgaagtgacacgtctcaagt 288
QY 421 caagtgcagttcagaaccatctgctctcagaaggagccaatactgaacaagaac 480
Db 289 caagtgcagttcagaaccatctgctctcagaaggagccaatactgaacaagaac 348
QY 481 cagtcctctctcacagccttgatgatatccttctactactagctctctgcctct 540
Db 349 cagtcctctctcacagccttgatgatatccttctactactagctctctgcctct 408

QY 541 gaaaatgcaggtagaccattcaaaactctgtcttaaccctctgcatactatacatcc-ac 599
Db 409 gaaaatgcaggtagaccattcaaaactctgtcttaaccctctgcatactatacatccnac 468
QY 600 cagtgcagctgcagaaagcataaaccctactgtagaactaaatgcactgtgcatgaact 659
Db 469 cagtgcagntgcagaaagcataaaccctactgtagaactaaatgccttggcatgaact 528
QY 660 tggaaaaaaaccaatgtataagcc-tgttgacccttactctc-ggatgcagtcacactat 717
Db 529 ggaaaaaaaccaaatgtataagccntgttgaaacttactctcggatgcagncacactat 588
QY 718 aactacaaca 727
Db 589 aactacaaca 598

Search completed: April 5, 2001, 08:31:19
Job time: 59323 sec

153 CGCAATGCAGCCGAGAACATGCTGAGATCCTGTTTCAAGTCCGCA 202
358 nArgGlnProThrIysProAlaLeuLysSerGlnGluLysThrProIleL 375
203 GCGGCAGCCACCAACCCGCACTCAAGTCAGAGAGAGAACACCCATAA 252
375 yslYsProGlyAspGlyArgLysValThrPhePheGluProGlySerGly 391
253 AGAAGACAGGGGATGAGAGAAAGTAACCTTTTGAACCTGGCTGGG 302
392 AspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetProTyrLe 408
303 GATGAAATGGAGCTAGTAATAAGAGAGATGAGTTCAGGATGCCCTTATCT 352
408 userHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluValA 425
353 AAGTCATCAGCAGCTGCCCTGCTGGAATTCCTCCATGCTGCCGAGTCCG 402
425 IacGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArg 441
403 CCCAGGCTGTAGAGTTAGTCAAGACATCACCAAGATTTTACCAGG 452
442 AlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgL 458
453 GCAGCTCCGAATCCCTGCCAAGGCCACGGTAACTGCCATGATAGCCGAGA 502
458 uLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnA 475
503 GTTGTGCTATGGGGGACCTCGCCACAGCCGAGACCATTTTAAAGATA 552
475 snIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 491
553 ACATCTCTTCAGGCCACGTACCCCATGACCTCTCACGAGACCCCTGAG 602
492 GlnLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAs 508
603 CAACTGACTATCTTCCAGAGTCCAGGAGTTCAGGTTGAATACAAAGA 652
508 PheProLysAsnAsn.LysAsnGluPhe.ValSer.LeuIleAsnCysS 524
653 CTTCGCCAAATAATCACCAAGACGAATTCGTAATCTCTTATCAATGCGCT 702
524 erSerGlnProProLeuIleSerHisGlyIle.GlyLysAspValGluSe 540
703 CCTCTCAGCCACT.CTGATCCACACATGGTATTCGGCAGGTTGTGGAAATC 751
540 rCysHis..AspMetAlaAlaLeuAsnIleLeuLysLeu.LeuSerGlu 555
752 CTGGCAATGATCATGGGTGGCTGGACATCTTTAAGATGGCTGGCGAA 801
556 leuAspGlnGlnSerThr.GluMetProArgThrGlyAsnGlyPromets 572
802 GTGACCCACCAAGGTCAAAAAATTCGCCAGAAACGGGAGGAGCCCACTT 851
572 erVal 573
852 CGGTG 856

seq_name: gb_est77:BE894295

seq_documentation_block:

LOCUS BE894295 752 bp mRNA EST 29-SEP-2000
DEFINITION 601437666F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922529 5',

mRNA sequence.

ACCESSION BE894295
VERSION BE894295.1 GT:10356518

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 752)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
RISSE Procurement: ATCC/DCTD/DTU
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM744 row: d column: 18
High quality sequence stop: 654.

FEATURES
source location/Qualifiers
1..752

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922529"
/clone_lib="NIH_MGC_72"
/issue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 223 a 194 c 192 g 143 t
ORIGIN

alignment_scores:
Quality: 1036.50 Length: 249
Ratio: 4.468 Gaps: 4
Percent Similarity: 93.173 Percent Identity: 89.157

alignment_block:
US-09-316-048-6 x BE894295 ..

Align seg 1/1 to: BE894295 from: 1 to: 752

281 TyrGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGlnAla 297
3 TATGGCCAGGGGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGGC 52
297 alYsLysGluLysGluProGluTyrThrLeuLeuThrGluArgGlyLeuP 314
53 AAAAAAGAGAGAGAGCCAGAGTACACGCTCTCCACAGAGCGAGGCTCC 102
314 roArgArgArgGluPheValMetGlnValLysValGlyAsnHisThrAla 330
103 CGCGCCGAGGAGATTGTGATGACAGGTGAAGTTGGAACCACACTGCA 152
331 GluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaIaGluAs 347
153 GAAGGAACGGGCACCAACAAGAGGTGGCCAAGCGCAATGCAGCGAGAA 202
347 nMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProThrLysP 364
203 CATGCTGAGATCCTTGGTTTCAAGTCCCGCAGCGCAGCCACCAAC 252
364 roAlaLeuLysSerGlnGluLysThrProIleLysLysProGlyAspGly 380
253 CCGCACTCAAGTCAGAGAGAGAGACACCCATAAGAAACCAAG.GATGGA 301
381 ArgLysValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSe 397
302 AGAAAGTAACCTTTTGAACCTGGCTGTGGGATGAAATGGGACTAG 351
397 rAsnLysGluAspGluPheArgMetProTyrLeuSerHisGlnGlnLeuP 414
352 TAATAAAGAGATGAGTTCAGGATGCCCTTATCTAAGTCATCAGCAGCTGC 401

414 roAlaGlyIleLeuPrometValProGluValAlaGlnAlaValGlyVal 430
|||||
402 CTGCTGGAATTCCTCCCATGGTCCCGAGGTCCGCCAGGCTGTAGAGATT 451
431 SerGlnGlyHisHisThrLysAspPheThrArgAlaAlaProAsnProAl 447
|||||
452 AGTCAGAGACATCACACCAAGATTTTACCAGACAGCTCCGAATCCTGC 501
447 aLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyLys 464
|||||
502 CAAGGCCACGTAAGTACGTAGATGATGATGATGATGATGATGATGATGAT 551
464 hrSerProThrAlaGluThrIleLeuLys.AsnaSnIleSerSerGlyHis 480
|||||
552 CCGGCCACACAGCCGAGATCATTTTAACGAATACATCTCTCAGGCCA 601
480 sValProHisGlyProLeuThrArgProSerGluGln.LeuAspTyrLeu 496
|||||
602 CGTACCCCATGGA..CTCTCAGCAGATCCTCTGAGCAATGACTATCTT 648
497 SerArgValGlnGlyPheGlnValGluTyrLysAspPheProLysAsna 513
|||||
649 TCCAGAGTCCA.GGATTCGCGGTGATACAAAGACTTCCCAAAAAACA 697
513 nLysAsnGluPheValSerLeuIleAsnCySerSerGlnPro 527
:|||||
698 GAAGAAATGGTTC.....TCTATCAATGCTCTCAACACCT 734

seq_name: gb_est74:BE674157

seq_documentation_block:
LOCUS BE674157 616 bp mRNA EST 08-SEP-2000
DEFINITION 7476a04.x1 NCI_CGAP_Lu24 Homo sapiens CDNA clone IMAGE:3278862 3'
similar to TR:095793 095793 STAU6N PROTEIN. ; mRNA sequence.

ACCESSION BE674157
VERSION BE674157.1 GI:10034698
KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 616)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 469.

FEATURES Location/Qualifiers

1. 616

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3278862"

/clone_lib="NCI_CGAP_Lu24"

/tissue_type="carcinoid"

/lab_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized

library NCI_CGAP_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was

used as tracer in a subtractive hybridization reaction.

The driver was PCR-amplified cDNAs from a pool of 5,000

clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT 179 a 153 c 139 g 141 t 4 others

ORIGIN

alignment_scores:

Quality: 1019.00

Ratio: 5.146

Percent Similarity: 97.537

Percent Identity: 96.059

alignment_block:

US-09-316-048-6 x BE674157 ..

Align seg 1/1 to: BE674157 from: 1 to: 616

371 LysThrProIleLysLysProGlyAspGlyArgLysValThrPhePheG 387
|||||
8 AGACACCCCATTAAGAACCACGAGGAGTGAAGAAAGTAACCTTTTGTGA 57
387 uProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspGluPhe 404
|||||
58 ACCTGGCTCTGGGGATGAATAATGGGACTAGTAATTAAGAGAGATGATTCA 107
404 rGmetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPromet 420
|||||
108 GGATGCTTATCTAAGTCATCAGCAGCTGCTGCGAATCTTCCCATG 157
421 ValProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrly 437
|||||
158 GTGCCCGAGGTCCGCCAGGCTGAGAGTAGTCAAGACATCACACCAA 207
437 sAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAla 454
|||||
208 AGATTTTACCAAGGCGAGCTCCGAATCCTGCCAAGGCCAGGTAAGTCCA 257
454 etIleAlaArgGluLeuLeuTyrGlyThrSerProThrAlaGluThr 470
|||||
258 TGATAGCCCGAGAGTGTGTGTATGGGGGCACCTCGCCACACCGAGACC 307
471 IleLeuLysAsnAsnIleSerSerGlyHisValProHisGlyProLeuth 487
|||||
308 ATTTTAAGATAATACATCTCTTCAGGCCACGTAACCCCATGGAACCTCTCAC 357
487 rArgProSerGluGlnLeuAspTyrLeuSerArgValGlnGlyPheGln 504
|||||
358 GAGACCCCTTGAGCAACTGACTATCTTCCAGAGTCCAGGATTCACAG 407
504 aGluTyrLysAspPheProLysAsnLysAsnGluPheValSerLeu 520
|||||
408 TTGAATACAAAGACTTCCCAAAACAACAAGACGAATTTGTATCTCTT 457
521 IleAsnCySerSerGlnProProLeuIleSerHisGlyIleGlyLysAs 537
|||||
458 ATCAATTGCTCTCTCANCCACCTCTGATCAGCCATGGTATCAGCAAGA 507
537 pValGluSerCyHisAspMetAlaAlaLeuAsnIleLeuLysLeuLys 554
|||||
508 TGTGAGTCTGCGCATGATATGCTGCGCTGAACATCTTAGAGTTGCTGT 557
554 erGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyPro 570
|||||
558 CTGAATGTGACCAACAANATCAGAGATGCCAAGAAGCANAGACAGACCA 607
571 MetSerVal 573
|||||
608 ATGCTCTGTA 616

seq_name: gb_est73:BE613972

seq_documentation_block:

LOCUS BE613972 686 bp mRNA EST 24-AUG-2000

/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAGAGATCCCTTAATTAATTAATCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTCGAGTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."

BASE COUNT 130 a 152 c 123 g 188 t 1 others
ORIGIN

alignment_scores:
Quality: 1004.00 Length: 198
Ratio: 5.122 Gaps: 0
Percent Similarity: 98.990 Percent Identity: 98.485

alignment_block:
US-09-316-048-6 x AW162002/rev ..

Align seg 1/1 to reverse of: AW162002 from: 1 to: 594

69 GluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysLeuG1 85
|||||
594 GAAAGCATACCCCTACTGTAGAACTANATGCACCTGTCATGAACCTTGG 545
85 YLysLysProMetTyrLysProValasProTyrSerArgMetGlnSert 102
|||||
544 AAAAAACCAATGTATAAGCCTGTGACCCCTTACTCTCGATGCAGTCCA 495
102 hTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProArgTyrPheTyr 118
|||||
494 CCTATACTACAACATGAGAGAGAGTGTATCCCGAGGTACTTTTAC 445
119 ProPheProValProProLeuLeuTyrGlnValGluLeuSerValGlyG1 135
|||||
444 CCATTTCAGTCCACCTTACTTATCAAGTGAACCTTCTGTGGAGG 395
135 YGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspa 152
|||||
394 ACAGCAATTATATGGCAAGAAAGACACAGAGGCTGCCAAGACGATG 345
152 laAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArg 168
|||||
344 CTGCTGCCAAGCGTTGAGGATCTGCAGAAATGAGCCCTGCCAGAGAGG 295
169 LeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSerg1 185
|||||
294 CTGAGGTGAATGGAAGAAATCCGAAGAGAAAAATCTCAATAAATCTGA 245
185 uIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsnP 202
|||||
244 AATAAGTCAAGTCTTTGAGATTCACACTTAACGGAACCTGCCCTGGAATT 195
202 heGluValAlaArgGluSerGlyProProHisMetLysAsnPheValThr 218
|||||
194 TCGAGGTGGCCCGGAGAGAGTGGCCACCCACATGAAGAACTTGTGACC 145
219 LysValSerValGlyGluPheValGlyGluGlyLysSerLysLy 235
|||||
144 AAGCTTCGTTGGGGAGGTTTGTGGGGGAGGTGAAGGAAACGACAGAA 95
235 sIleSerLysLysAsnAlaAlaIleAlaValLeuGluLeuLysLysL 252
|||||
94 GATTTCAAAGAAAAATGCCGCATAGCTGTTCTTGAGGAGCTGAAGAAAGT 45

252 euProProLeuProAlaValGluArgValLysProArgIleLys 266
|||||
44 TACCGCCCTGCCCTGCAGTTGAACGAGTAAGCCTAGAAAAAAA 1

seq_name: gb_est69:BE300155

seq_documentation_block:
LOCUS BE300155 829 bp mRNA EST 20-JUL-2000
DEFINITION 600944602T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960593 3',
mRNA sequence.
ACCESSION BE300155
VERSION BE300155.1 GI:9183903
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LTCM54 row: d column: 02
High quality sequence start: 41
High quality sequence stop: 799.
Location/Qualifiers
1. 829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960593"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOT87; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

FEATURES
source

BASE COUNT 190 a 200 c 226 g 213 t
ORIGIN

alignment_scores:
Quality: 947.00 Length: 217
Ratio: 4.832 Gaps: 2
Percent Similarity: 90.323 Percent Identity: 87.097

alignment_block:
US-09-316-048-6 x BE300155/rev ..

Align seg 1/1 to reverse of: BE300155 from: 1 to: 829

36 ProSerThrThrSerSerLeuProSerGluAsn.....AlaGlyArgPr 50
|||:::||||:::|||||
636 CCACCTTAACCTCTCAGAACTGAACAAGACACAACTGTTCCTGGACGCC 587
50 oIleGlnAsnSerAlaLeuProSerAlaSerIleThrSerThrSerAla 67
|||:::|||||
586 TCTTTTAAAAAA..... 574
67 laAlaGluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLys 83

|||||
573GAAGCATACCCCTACTGTAGAAGTAATGCACCTGTGCATGA 529
84 LeuGlyLysLysProMetTyrLysProValAspProTyrSerArgMetG1 100
528 CTTGGAAAAAACCAATGATATAAGCCTGTGACCCCTTACTCTCGGATGCA 479
100 nSerThrTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProArgTyrP 117
478 GTCCACCTATACTACACATGAGAGAGAGGTGCTTATCCCCGAGTACT 429
117 heTyrProPheProValProProLeuLeuTyrGlnValGluLeuSerVal 133
428 TTTACCCATTTCAGTTCACCTTACTTATCAAGTGGAACCTTCTGTG 379
134 GlyGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysH1 150
378 GGAAGACAGCAATTTAATGGCAAGAAAGAACAGACAGGCTGCGAACA 329
150 sasPaAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProG 167
328 CGATGCTGCTGCCAAGCGTTGAGGATCCTGCAATGAGCCCCCTGCCAG 279
167 luArgLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLys 183
278 AGAGGCTGGAGGTGAATGGAAGAAATCCGAGACAGAAATCTCAATAAA 229
184 SerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProVa 200
228 TCTGAATAATAGTCAAGTGTGATGATTGACATTAAACGGAACCTTGCTGT 179
200 lasnPheGluValAlaArgGluSerGlyProProHisMetLysAsnPhy 217
178 GAATTTGAGGTGGCCGGGAGAGTGGCCACCCACATGAAGAATTTTG 129
217 alThrLysValSerValGlyLupheValGlyGluGlyLysSer 233
128 TGACCAAGGTTTCGTTGGGAGAGTTTGTGGGGAAGGTGAAGGGAAGAC 79
234 LysLysIleSerLysLysAsnAlaAlaIleAlaValLeuGluLeuLys 250
78 AAGAAGATTTCAGAAAGAAATGCCCATAGCTGTCTGAGAGAGCTGAA 29
250 s 250
28 G 28
seq_name: gb_est43:AW320695
seq_documentation_block:
LOCUS AW320695 633 bp mRNA EST 25-JAN-2000
DEFINITION uo2b10.y1 NCI-CGAP_Mam6 Mus musculus cDNA clone IMAGE:2609851 5'
similar to TR:O95793 O95793 STAUFEIN. ; mRNA sequence.
ACCESSION AW320695
VERSION AW320695.1 GI:6750239
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 633)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
MGI:1023711
Seq primer: -40RP from Gibco
High quality sequence stop: 356.
location/Qualifiers
1. 633
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:2609851"
/clone_lib="NCI-CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pcMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 176 a 167 c 163 g 127 t
ORIGIN
alignment_scores:
Quality: 941.00 Length: 206
Ratio: 4.801 Gaps: 0
Percent Similarity: 95.146 Percent Identity: 86.893
alignment_block:
US-09-316-048-6 x AW320695 ..
Align seg 1/1 to: AW320695 from: 1 to: 633
358 GlnArgGlnProThrLysProAlaLeuLysSerGluGluLysThrProil 374
16 CAGGCGCAGCGCTGCCAAGCCAGCACACTCAATATCAGAAGAGAGACTCCAGT 65
374 elysLysProGlyAspGlyArgLysValThrPhePheGluProGlySerg 391
66 AAGAAGAACAGAGAGAGAGAGAAAGTAACGTTTGTGAACCTAGCCCTG 115
391 lYAspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetProtyr 407
116 GGCATGAAAATGGAAGTACTAACAAGAGAGAGAGATTCAGATGCCCTTAT 165
408 LeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluVa 424
166 CTTAGCCATCAGCAGCTGCCAGCTGGAATTCCTCCATGTGTCGGGAGAT 215
424 lAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThra 441
216 TGCCCAAGGCTGTGGGGCTTAGTCAAGGACACACCAAGATTTCACCA 265
441 rgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 457
266 GGGCAGCTCCAAATCTCGCCCAAGGCAACGGTAACTGCCATGATAGCCCGA 315
458 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAs 474
316 GAGTTGTTGTACGGGGGACCTCGCCACAGCCGAGACCATTTTAAAGAG 365
474 nasnIleSerSerGlyHisValProHisGlyProLeuThrArgProSerg 491
366 TAACATCTCTTACAGCCACGTACCCGATGGACCTCGCACTAGACCCCTG 415
491 lueGlnLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLys 507
416 AGCAACTGTACTACTTCCACAGAGCCAGGAGTTCAGGTGAATACAA 465
508 AspPheProLysAsnAsnLysAsnGluPheValSerLeuIleAsnCysSe 524
|||||

```
466 GATGTTGCCAAGACGACGACGAGTGTGTATCTCTCATCACTGCTC 515
524 rSerGlnProProLeuIleSerHisGlyIleGlyLysAspValGluSerC 541
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
516 CTCACAGCCGGCTCTCGTCAGTCATGGCATCGCGCAAGATGTGAAGTCC 565
541 ySHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLeuAsp 557
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
566 GTCATGATATATGGCTGCACCTGCATTTGTAAGCTGCTGTGTGAGTTGAC 615
558 GlnGlnSerThrGluMet 563
:::|||||:|||||:
616 TCACAGAACACAGATATT 633

seq_name: gb_est66:BE082712

seq_documentation_block:
LOCUS      BE082712      615 bp      mRNA      EST      12-JUN-2000
DEFINITION RC2-BT0642-150200-012-a08 BT0642 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE082712
VERSION    BE082712.1  GI:8473017
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 615)
AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE     20202663
COMMENT     Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=RC2-BT0642-150
             200-012-a08&ct3=2000-02-15&ct4=1)
             Seq primer: puc 18 forward
             High quality sequence start: 75
             High quality sequence stop: 609.
FEATURES
source
1..615
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BT0642"
    /dev_stage="Adult"
    /note="Organ: breast; Vector: puc18; Site.1: SmaI; Site.2:
    SmaI; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    ,716 - Ludwig Institute for Cancer Research) profiles
    into the pUC 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."
BASE COUNT  110 a      159 c      153 g      193 t
ORIGIN
alignment_scores:
Quality:      936.00      Length:      187
Ratio:        5.032      Gaps:        0
Percent Similarity: 99.465      Percent Identity: 97.326
```

```
alignment_block:
US-09-316-048-6 x BE082712/rev ..
Align seg 1/1 to reverse of: BE082712 from: 1 to: 615

232 LysSerLysLysIleSerLysLysAsnAlaAlaIleAlaValLeuGlu 248
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
613 AAAAGCAGAAGATTTCAAAGAAAATGCCCATAGCTGTTCTTGAGGA 564
248 uLeuLysLysLeuProProLeuProAlaValGluArgValLysProArgI 265
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
563 GCTGAAGAAGTTACCGCCCTGCTGGAGTAGAACGAGTAAGCCTAGAA 514
265 lLysLysLysThrLysProIleValLysProGlnThrSerProGluTyr 281
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
513 TCAAAAAGAAAACCAAAACCATAGTCAAGCCACAGACAGACCCAGAATAT 464
282 GlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGlnAlaLys 298
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
463 GGCCAGGGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGCAAA 414
298 sLysGluLysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProA 315
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
413 AAAGAGAGAGAGAGCCAGAGTAGACGCTCCTCACAGAGCAGGCTCCCGC 364
315 rGArgArgGluPheValMetGlnValLysValGlyAsnHisThrAlaGlu 331
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
363 GCCGACAGGAGTTGTGATGCAGGTGAAGTTGGAACCAACACACTGCAGAA 314
332 GlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMe 348
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
313 GGAACGGGCAACAACAAGAAGGTGGCCAAGCCCAATGCAGCCGAGAACAT 264
348 tLeuGluIleLeuGlyPheLysValProGlnArgGlnProThrLysProA 365
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
263 GCTGAGATCCTTGTTTCAAGTCCCGCAGCGCAGCCACCAAAACCGG 214
365 lAlaLysSerGluGluLysThrProIleLysLysProGlyAspGlyArg 381
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
213 CACTCAAGTCAGAGAGAGAACACCATAAAGAAACAGGGGATGGAAGA 164
382 LysValThrPheGluProGlySerGlyAspGluAsnGlyThrSerAs 398
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
163 AAAGTAACCTTTTGAACCTGCTGGGATGAATAATGGGACTAGTAA 114
398 nLysGluAspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProA 415
|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
113 TAAAGAGGATGAGTTCAGGATGCCTTATCTAAGTCATCACCATCTTCCTG 64
415 laGlyIleLeu 418
|||||  |||:|||||
63 CTGACCTTCTT 53

seq_name: gb_est40:AW163206

seq_documentation_block:
LOCUS      AW163206      564 bp      mRNA      EST      09-NOV-1999
DEFINITION au97f05.x1 schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2784225 3' similar to TR:095793 095793 STAUFEN PROTEIN. ;
MRA sequence.
ACCESSION  AW163206
VERSION    AW163206.1  GI:6302239
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 564)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
             Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maira,M., Martin
             ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
             White,Y., Wylie,T., Waterston,R. and Wilson,R.
```

TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 455.

FEATURES
source
1. .564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2784225"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGAGCTCAAGATCCTTAATTAATTAATCCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGAGCTCGAGTTTCTTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."

BASE COUNT 125 a 147 c 117 g 175 t
ORIGIN

alignment_scores:
Quality: 926.00 Length: 188
Ratio: 5.060 Gaps: 0
Percent Similarity: 97.340 Percent Identity: 96.277

alignment_block:
US-09-316-048-6 x AW163206/rev ..

Align seg 1/1 to reverse of: AW163206 from: 1 to: 564

79 AlaleuCysMetLysLeuGlyLysLysProMetTyrrLysProValaspPr 95
|||||::: ||||||||| |||||
564 GCACTGTGCATAACACTTGGAAAAAACCAATGTATTAGCCTGTGACCT 515
95 oTyrrSerArgmetGlnSerThrTyrrAsnTyrrAsnMetArgGlyAlaT 112
||||| ||||||||| |||||::: |||||||||
514 TTACTCTCGATGACATCCACCTATACTACACATGAGAGAGGTGCTT 465
112 yrrProPArgTyrPheTyrrProPheProValProProLeuLeuTyrrGln 128
||||| ||||||||| ||||||||| |||||||||
464 ATCCCCCGAGGTACTTTTACCCATTTCACGTTCCACCTTACTTTATCAA 415
129 ValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLysGlyLysThrAr 145
||||| ||||||||| ||||||||| |||||||||
414 GTGGAACCTTCTGTGTGGAGGAGACAGCAATTATGGCAAGAAAGACAAG 365
145 gGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuArgIleLeuGlnA 162
||||| ||||||||| ||||||||| |||||||||
364 ACAGGCTGCGAAACAGATGCTGTGCCAAGCGTTGAGGATCCTGCAGA 315
162 sngLuproLeuProGluArgLeuGluValAsnGlyArgGluSerGluGlu 178
||||| ||||||||| ||||||||| |||||||||
314 ATGAGCCCTGTCCAGAGAGGCTGGAGGTGAATGGAAGAATCCGAGAA 265

179 GluAsnLeuAsnLysSerGluIleSerGlnValPheGluIleAlaLeuLy 195
||||| ||||||||| ||||||||| |||||||||
264 GAAATCTCAATTAATCTGAATAAGTCAAGTGTTCAGATTGCACCTTAA 215
195 sArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProh 212
||||| ||||||||| ||||||||| |||||||||
214 ACGGAACCTTGCCCTGTGAATTTTCAGAGGTGGCCCCGGAGAGTGGCCCC 165
212 IsMetLysAsnPheValThrLysValSerValGlyGluPheValGlyGlu 228
||||| ||||||||| ||||||||| |||||||||
164 ACATGAAGAACTTGTGTGACCAAGGTTTCGTTGGGGAGTTTGTGGGGAA 115
229 GlyGluGlyLysSerLysLysIleSerLysLysAsnAlaAlaIleAlaVa 245
||||| ||||||||| ||||||||| |||||||||
114 GGTGAAGGAGAAAGCAAGATTTCAAAAGAAATGCCGCATAGCTGT 65
245 IleuGluGluLeuLysLysLeuProProLeuProAlaValGluArgYall 262
||||| ||||||||| ||||||||| |||||||||
64 TCCTTGAGGAGAGCTGAAGAGTTACCGCCCTGCCTGCAGTTGAACGAGTAA 15
262 ysrProArgIleLys 266
||||| |||
14 AGCCTAGAAAAAAA 1

seq_name: gb_est73:BE617223

seq_documentation_block:
LOCUS BE617223 749 bp mRNA EST 24-AUG-2000
DEFINITION G01441906F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846265 5',
mRNA sequence.
ACCESSION BE617223
VERSION BE617223.1 GI:9888161
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 749)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM545 row: k column: 02
High quality sequence stop: 705.

FEATURES
source
1. .749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846265"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 244 a 166 c 223 g 116 t
ORIGIN

alignment_scores:
Quality: 925.00 Length: 254
Ratio: 3.903 Gaps: 7

Percent Similarity: 93.307 Percent Identity: 90.551

alignment_block:

US-09-316-048-6 x BE617223 ..

Align seg 1/1 to: BE617223 from: 1 to: 749

```
143 LysThrArgGlnAlaAlaLysHisAspAlaAlaLysAlaLeuArgI 159
|||||
1 AAGACAAGACAGAGCTGCCAACAACGATGCTGCTGCCAAGGCGTTGAGGA 50
159 lEleuGlnAsnGluProLeuProGluArgLeuGluValAsnGlyArgGlu 175
|||||
51 TCCTGCAGAATGAGCCCTGCCAGAGAGGCTGAGGTGAATGAAGAGAA 100
176 SerGluGluGluAsnLeuAsnLysSerGluIleSerGlnValPhe.GluI 192
|||||
101 TCCGAAGAAGAAATCTCAATAAATCTGAATAATGCAAGTGTGGAGA 150
192 leAlaLeuLysArgAsnLeuProValAsnPheGluValAlaAArgGluSer 208
|||||
151 TTGCACTTAACGGAAGACTTGCTGTGAATTTGAGGTGGCCCGGAGAGT 200
209 GlyProProHisMetLysAsnPheValThrLysValSerValGlyGluPh 225
|||||
201 GGGCCACCCACATGAAGAACTT.GTGACCAAGTTTCGGTGGGAGTT 248
225 eValGlyGluGlyGlyLysSerLysLysIleSerLysLysAsnAlaA 242
|||||
249 .GTGGGGGAAGGTGAAGGGAACAAGAGATTTCAAAGAAAATGCGG 298
242 laIleAlaValLeuGluGluLeuLysLysLeuProProLeuProAlaVal 258
|||||
299 CCATAGCTGTCTTGTAGAGAGCTGAAGAAGTTACCCTCTG.CCTGCAGTT 347
259 GluArgValLysProArgIleLysLysLysThrLysProIleValLysPr 275
|||||
348 GAACGAGTAAGCCTAGAAATCAAAAAGAAACAACCAACCATAGTCAAGCC 397
275 oGlnThrSerProGluTyrGlyGlnGlyIleAsnProIleSerArgLeuA 292
|||||
398 ACAGACAAGCCCAAGATATGGCCAGGATC.AATCCGATTAGCCGACTGG 446
292 laGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrThrLeuLeu 308
|||||
447 CCCAGATCCAGCAGGCAAAAAAGAGAAGAGCCAGAGTACACGCTCTTC 496
309 ThrGluArgGlyLeuProArgArgArgGluPheValMetGlnValLysVa 325
|||||
497 ACAGAGCGAGGCTC.CCGCGCCGACAGGAGTT.GTGATGCAGGTGAAGGT 544
325 lGlyAsnHisThrAlaGluGlyThrGlyThrAsn.LysLysValAlaLys 341
|||||
545 GGGAAACCACTGACAGAAGAACGGGCACCAACAAGAAGGTGGCCAG 594
342 ArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPhe..LysValPro 357
|||||
595 CGCAATGCAGCCGAGACATGCTGGAGATCCTTGCTTCAAAAGTCCCG 644
358 GlnArgGlnProThrLysProAlaLeuLysSerGluLysLysThrProI 374
|||||
645 CAGGGGACGCCAACAAAC...CGGACTCAGTCAGAGGAGAAGACACCCAT 691
374 eLysLys.ProGlyAspGlyArgLysValThrPhePheGluProGlySer 390
|||||
692 A...AGAAACCGGGGATGGAAGACACGTAAACCTTTTGAACCTGGGCGG 738
391 Gly 391
|||
739 GGA 741
```

seq_name: gb_est75:BE781223

seq_documentation_block:

LOCUS BE781223 673 bp mRNA

EST 20-SEP-2000

DEFINITION 601469159F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3872247 5', mRNA sequence.

ACCESSION BE781223

VERSION BE781223.1 GI:10202421

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM613 row: e column: 16

High quality sequence stop: 644.

location/Qualifiers

1. 673

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3872247"

/clone_lib="NIH_MGC_67"

/tissue_type="retinoblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Library constructed by Life

Technologies."

BASE COUNT 241 a 156 c 178 g 98 t

ORIGIN

alignment_scores:

Quality: 908.50 Length: 227

Ratio: 4.635 Gaps: 4

Percent Similarity: 86.344 Percent Identity: 85.022

alignment_block:

US-09-316-048-6 x BE781223 ..

Align seg 1/1 to: BE781223 from: 1 to: 673

```
182 AsnLysSerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLe 198
|||||
1 AATAAATCTGAAATAGTCAAGTGTGTGAGATTGCACCTTAACGGAAGT 50
198 uproValAsnPheGluValAlaAArgGluSerGlyProProHisMetLysA 215
|||||
51 GCCTGTGAATTTGAGGTGGCGCCGAGAGAGTGCCCAACCCACATGAAGA 100
215 snPheValThrLysValSerValGlyGluPheValGlyGlyGlyGly 231
|||||
101 ACTTGTGACCAAGGTTTCGGTGGGAGTTTGTGGGGGAAGTGAAGG 150
232 LysSerLysLysIleSerLysLysAsnAlaAlaIleAlaValLeuGlu 248
|||||
151 AAAAGCAAGAGATTTCAAAGAAATAATGCCCATAGCTGTTCTTGAGGA 200
248 uLeuLysLysLeuProProLeuProAlaValGluArgValLysProArgI 265
|||||
201 GCTGAAGAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAGCCTAGAA 250
```

```
265 leLysLysLysThrLysProIleValLysProGlnThrSerProGluTyr 281
|||||
251 TCATAAAGAAAAACAAACCCATAGTCAAGCCACAGACAGCCAGCATAT 300
282 GLyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGlnAlaIy 298
|||||
301 GGCCAGGGGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGCAAA 350
298 sLySGluLysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProA 315
|||||
351 AAAGAGAAGAGAGCCAGATACACGCTCCTCACAGAGCGAGCCCTCCCGC 400
315 rgArgArgGluPheValMetGlnValLysValGlyAsnHisThrAlaGlu 331
|||||
401 GCCCGAGGAGTGTGTGATGCAGGTGAAGGTTGGAACACACACTGCAGAA 450
332 GLyThrGLyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsn 348
|||||
451 GGAACGGGCACCAACAAGAGTGGCCAAGCGCAATGCAGCCGAGACAT 500
348 tLeuGluIleLeuGly.PheLysValProGlnArgGlnProThrLys..P 364
|||||
501 GCTGGAGATCCTTGCTTTCAAAGTCCCGCAGCGCCAGCCACCAAAACC 550
364 roAlaLeuLysSerGluGluLysThrProIleLys..... 375
|||||
551 CAAGACACCCCCCGAAAAAATAAAGAAGAAACATATAAATAA 600
376 .....LysProGlyAspGlyArgLy 382
|||||
601 AACACAATTAACACCAACATTAACAACAAAAAAGGGGGGGGGCCCC 650
382 sValThrPheGluProGlySerGly 391
|||||
651 AAAAAAGTTC.....CCCGGGGGGGC 672
```

seq_name: gb_est3:AA206573

seq_documentation_block:

LOCUS AA206573 573 bp mRNA EST 12-MAR-1998
DEFINITION zq51f08.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
clone IMAGE:645159 5' similar to SW:STAU_DROME P25159 MATERNAL
EFFECT PROTEIN STAUFEN. ; mRNA sequence.

ACCESSION AA206573

VERSION AA206573.1 GI:1801953

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 573)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Mairra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.

AUTHORS

TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

FEATURES

source This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 3237 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 325.

FEATURES

1.573
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:5215410"
/db_xref="taxon:9606"

```
/clone="IMAGE:645159"  
/clone_lib="Stratagene neuroepithelium (#937231)"  
/dev_stage="Ntera-2/Ra neuroepithelial cells"  
/lab_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2  
cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24  
hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5'  
adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor  
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"  
BASE COUNT 178 a 118 c 145 g 129 t 3 others  
ORIGIN
```

```
alignment_scores:  
Quality: 882.00 Length: 191  
Ratio: 4.642 Gaps: 0  
Percent Similarity: 99.476 Percent Identity: 96.859
```

alignment_block:
US-09-316-048-6 x AA206573 ..

Align seg 1/1 to: AA206573 from: 1 to: 573

```
69 GluSerIleThrProThrValGluLeuAsnAlaLeuGluCysMetLysLeuG 85  
|||||  
4 GAAGCATTAACCCCTACTGTAGACTAATGCACCTGTGCATGAACCTTG 53
```

```
85 yLysLysProMetLysLysProValAspProTyrSerArgMetGlnSert 102  
|||||  
54 AAAAAAACCAATGTATAAGCCTGTGACCCCTTACTCTCGGATGCACCTCA 103
```

```
102 hTYrAsnTYrAsnMetArgGlyGlyAlaTYrProProArgTYrPheTYr 118  
|||||  
104 CCTATACTACACATGAGAGAGGAGTGTATATCCCGGAGGACTTTTAC 153
```

```
119 ProPheProValProProLeuLeuTYrGlnValGluLeuSerValGlyG 135  
|||||  
154 CCATTTCAGGTTCCACCTTTACTTTATCAAGTGAACCTTCTGTGGGAGG 203
```

```
135 yGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspA 152  
|||||  
204 ACAGCAATTTAATGGCAAGAAAGAACAGACAGAGCTGCGAAACAGATG 253
```

```
152 IaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArg 168  
|||||  
254 CTGCTGCCAAGCGTTGAGGATCCTGCAGATGAGCCCTGCCAGAGAG 303
```

```
169 LeuGluValAsnGlyArgGluSerGluGluAsnLeuAsnLysSerG 185  
|||||  
304 CTGGAGGTGAATGGAAGAATCCGAAGAGAAATCTCAATTAATCTGA 353
```

```
185 uIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsn 202  
|||||  
354 AATTAAGTCAAGTGTGAGATTGCACCTTAACCGGAACCTTGCTGAAATT 403
```

```
202 heGluValAlaArgGluSerGlyProProHisMetLysAsnPheValThr 218  
|||||  
404 TCGAGGTGGCCCG. GAGAGTGGCCACCCACATGAAGAAGACTTGTGANC 452
```

```
219 LysValSerValGlyGluPheValGlyGluGlyGlyLysSerLysLy 235  
|||||  
453 AAGGTTTCGGTTGGGAG. TTTGTGGGGGAAGGTGAAGGAAAGCCAGAA 501
```

```
235 sIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLysLysL 252  
|||||  
502 GATTTC. AAGAAATAATGCCGCCATAGCTGTCTTGAGGAGGCTGAAGAAGT 550
```

```
252 euProProLeuProAlaValGlu 259  
::: |||||  
551 TTCNGCCCTGCCTGCNATTGAA 573
```

seq_name: gb_est77:BE872137


```
seq_documentation_block:
LOCUS      BE872137      852 bp      mRNA      27-SEP-2000
DEFINITION 601446238F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850496 5',
            mRNA sequence.
ACCESSION  BE872137
VERSION    BE872137.1  GI:10320913
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 852)
            NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
JOURNAL    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM556 row: k column: 09
            High quality sequence stop: 637.
            Location/Qualifiers
FEATURES
    source
        1..852
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3850496"
            /clone_lib="NIH_MGC_65"
            /tissue_type="adenocarcinoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.8 kb. Library constructed by Life
            Technologies."
BASE COUNT  275 a      215 c      218 g      144 t
ORIGIN
alignment_scores:
    Quality: 878.00      Length: 252
    Ratio: 4.028      Gaps: 9
    Percent Similarity: 86.508      Percent Identity: 78.175
alignment_block:
US-09-316-048-6 x BE872137 ..
Align seg 1/1 to: BE872137 from: 1 to: 852
295 gInGlnAlaLysLysGluLysGluProGluTyrThrLeuLeuThrGluAr
|||||
1 CAGCAGGCAAAAAGGAGAGAGCAGAGTACGCTCCTCACAGAGCG 50
311 gGlyLeuProArgArgGluPheValMetGlnValLysValGlyAsnH
|||||
51 AGGCCTCCCGCGCGCAGAGAGTTGTGATGACGTGAAGTTGGAACC 100
328 iSthraLagLuglyThrGlyThrAsnLysLysValAlaLysArgAsnAla 344
|||||
101 ACACCTGCAGAGAGACGGGACCAACAAGAGGTGCCCAAGCGCAATGCA 150
345 AlagLAsnMetLeuGluLeuLeuGlyPheLysValProGlnArgGlnPr
|||||
151 GCCGAGAACATGCTGTGAGATCCTTGTTCAAGTCCCGCAGCGCAGCC 200
361 oThrLysProAlaLeuLysSerGluGluLysThrProIleLysLysProG
|||||
201 CACCAAAACCGCACTCAAGTCAGAGAGAGAACACCAACCAACACAG 250
```

```
378 lYAspGlyArg...LysValThrPhePheGluProGlySerGlyAsp... 392
|||||
251 GGGATGGAAGAACACAGTCAACCTTTTGTGAACCTGGCTCTGGGGCATGC 300
393 GluAsnGlyThrSer.....AsnLysGluAspGluPhe.ArgMetPro. 406
:::
301 AAACATGGGCACCTCAGTATATCACAAAGCAGGCATGAGTTCAGGATGCC 350
407 ..TyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValPro 422
|||
351 CATCTAAGTCCATCCACAGCAGCTGCTGCTGGAATTCCTCCATGTGCC 400
423 GluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPh
|||||
401 GAGTCGCCAGAGCTGTGAGAGTTAGTCAAGGACATCACCAAGATTT 450
439 eThrArg...AlaAlaProAsnProAlaLysAlaThrValThrAlaMet. 454
|||||
451 TACCAGGACACAAGCTCCGACATCCTGACAAGGCCGCGTAACCTGACCTGC 500
455 .IleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrI 471
:::
501 ACTAGCCCGAGAGTGTGATATGGGGCACCCTGCCACACAGCCGAGACCA 550
471 lLeuLysAsnAsnIleSerSerGlyHisValProHisGlyProLeuThr 487
|||||
551 TTTTACAGATAATACATCTCTTCAGGCCACGTAACCATGGAACCTCTCACG 600
488 ArgProSerGluGlnLeuAspTyrLeuSerArgValGlnGlyPheGln.V 504
|||||
601 AGACCCTCTGAGCAACTGACTATCTTACCAGAGTCCAGGCATACCAAGG 650
504 alGluTyrLysAspPheProLysAsnAsnLysAsnGluPheValSer.Le 520
|||||
651 TTGACTACAAAGAGACTTCCCAAAAGCAAAAGACGAATAATATCTCTTT 700
520 uIleAsnCysSerSerGlnProProLeuIleSerHisGlyIleGlyLys 536
|||||
701 AATCAATTGTTACTCTCAGACAGCTCTGATCAGCAGAGAATCGGAAGG 749
seq_name: gb_est72:BE541462
seq_documentation_block:
LOCUS      BE541462      673 bp      mRNA      09-AUG-2000
DEFINITION 601067913F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3454179 5',
            mRNA sequence.
ACCESSION  BE541462
VERSION    BE541462.1  GI:9770107
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 673)
            NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
JOURNAL    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM8438 row: j column: 04
            High quality sequence stop: 637.
            Location/Qualifiers
FEATURES
    source
        1..673
            /organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:3454179"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT      208 a      168 c      183 g      114 t
ORIGIN

alignment_scores:
    Quality: 871.00      Length: 227
    Ratio: 4.167      Gaps: 7
Percent Similarity: 92.070      Percent Identity: 87.225

alignment_block:
US-09-316-048-6 x BE541462 ..

Align seg 1/1 to: BE541462 from: 1 to: 673

255 leuproalavalgluargvalysproargilelys.lyslysthrlysp 271
|||||
2 CTGCCTGCAGTTGAACGAGTAAGCCTAGAAATCAATAAGAAAAACAAC 51
|||||
271 rollevalysproglthrserproglutyrlyglnglyleasnpro 287
|||||
52 CCATAGTCAAGCCACAGACAGAGCCAGAAATATGGCCAGGGATCAATCCG 101
|||||
288 lleserargleualacnleaglinalalyslysglulysgluprogl 304
|||||
102 ATTAGCCGACTGGCCAGATCCAGCAGCAAAAAAGAGAGAGACCAGA 151
|||||
304 utyrthrleuleuthrgluargglyleuproargargargluPhevalm 321
|||||
152 GTACACGCTCCTCACAGAGAGGCGCTCCGCGCCGACGAGGAGTT.GTGA 200
|||||
321 etglvalysvalgllyasnhisthrslagluglythrlythrlysnlys 337
|||||
201 TGCAGGTGAAGGT.GGAAACCACTGCACAGGAAGGCGCACCAACAAG 249
|||||
338 lysvalalalysargasnalaalagllyasnmetleugluile.leuglyp 354
|||||
250 AAGGTGCCAAGCCCAATGCAGCCGAGAACATGCTGGAGATCCTTGGGTT 299
|||||
354 helysvalproglinarglprothrlys..proalaleuylsserglug 370
|||||
300 TCAAGTCCCGCAGCGCCAGCCCAACCCGACGCTCAAGTCAGAGG 349
|||||
370 lylsthrprolelelyslslyprogllyaspglyarglyvalthrphphe 386
|||||
350 AGAAGACACCCATAAAGAAACAGGGGATGGAAGAAAGTAACCTTTT 399
|||||
387 gluproglyserglyaspglu.asnglythrserasnlysgluaspglup 403
|||||
400 GAACCTGGCTCTGGGGATGAACATGGGACTAGTAATAAGAGAGATGAGT 449
|||||
403 heargmethrprolyrleuserhisglnleuproalagllyleuPro 419
|||||
450 TCAGGATGCCTTATCTAAGTCATCAGCAGCTGCTGGAATCTTCCC 499
|||||
420 metvalprogluvalalaglinalavalglyvalserglnlyhis.T 436
|||||
500 ATGGTGGCAGGTCGCCAGGCTTGTAGAGTTAGTCAAGGACATCACAT 549
|||||
436 hrlyaspphertharaglaalaproasnproalalys.alathrvalth 452
|||||
550 CCAAGATTTTACCAGGGCAGCTCCGAATCTGCCAAGGCGCACGGTAAC 599
|||||
452 ralametilealaargluLeuLeuTyrglyglythrserProthr.Ala 468
|||||
```

```
600 TGCATGATTAGCCGAGAGTTGTGTATGGGGCCACCTCGTCCAATAGCGG 649
469 gluThrIleuLysAsn 474
:::|||||
650 AAGACCATTTTACAGAAAT 667

seq_name: gb_est76:BE868328

seq_documentation_block:
LOCUS      BE868328      866 bp      mRNA      EST      27-SEP-2000
DEFINITION 601443887F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848252 5',
            mRNA sequence.
ACCESSION  .BE868328
VERSION    BE868328.1 GI:10317104
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 866)
AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM550 row: m column: 21
            High quality sequence stop: 605.

FEATURES             location/Qualifiers
     source           1..866
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3848252"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.8 kb. Library constructed by Life
                     Technologies."
BASE COUNT      288 a      229 c      224 g      125 t
ORIGIN

alignment_scores:
    Quality: 861.00      Length: 249
    Ratio: 4.005      Gaps: 3
Percent Similarity: 86.345      Percent Identity: 79.920

alignment_block:
US-09-316-048-6 x BE868328 ..

Align seg 1/1 to: BE868328 from: 1 to: 866

238 lyslysasnalaalealalealavalleugluLeuLysLysLeuProPr 254
|||||
5 AAAGAAAAATGCCGATAGCTGTTCTTGAGGAGAGCTGAAGAAGTTACCGCC 54
|||||
254 oleuproalavalgluargvalysproargilelyslyslslythrlysp 271
|||||
55 CCTGCCTGCAGTTGAACGAGTAAGCCTAGAAATCAAAAAGAAACAAC 104
|||||
271 rollevalysproglthrserproglutyrlyglnglyleasnpro 287
|||||
105 CATA.GTCAAGCCACAGACAGAGCCAGAAATATGGCCAGGATC.AATCCG 152
|||||
```

```

288 ILeSerArgLeuAlaGlnIleGlnGlnAla.LysLysGluLysGluProG 304
    |||
153 ATTAGCCGACTGGCCAGATCCAGCAGGCAAAAAAAGAGAGAGGCCAG 202
    |||
304 IuTyThrLeuLeuThrGluArgGlyLeuProArgArgGluPheVal 320
    |||
203 AGTACAGCCTCTCTACAGAGAGGCTCCCGCGCCGAGGAGTTGTA 252
    |||
321 .MetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnL 337
    |||
253 GATGCAAGTGAAGTTGGAACACACACTGCAGAGGAAGGCGCACCAACA 302
    |||
337 ySLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGly 353
    |||
303 AGAAGGTGGCCAAGCGCAATGCAGCCGAGACATGCTGGAGATCCTGGT 352
    |||
354 PheLysValProGlnArgGlnProThrLysProAlaLeuLysSerGluG 370
    |||
353 TTCAAAGTCCCCGAGGCGCAGCCCAACCAACCCGACTCAAGTCAGAGGA 402
    |||
370 uLysThrProIleLysLysProGlyAspGlyArgLysValThrPhePheG 387
    |||
403 GAAGACACCCATAAAGAAACCAAGGGGATGGAAAGAAAGTAACTTTTGG 452
    |||
387 IuProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspGluPhe 403
    |||
453 AACCTGGCTCTGGGATGAACATGGGACTAGTACTAAAGAGATGAGTTC 502
    |||
404 ArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProme 420
    |||
503 AGGATGCCCTTATCTAGTCATCAAGCAGCTGCCGTGGATCTTCCCAT 552
    |||
420 tValProGluValAlaGlnAlaValAlGlyValSerGlnGlyHisHisThrL 437
    |||
553 GGTGACCGCAGCTGCG.CAGGCTGTAGGAGTCAGTCAAGGACATCACACCA 601
    |||
437 ySAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAla 453
    |||
602 CAGATTA.ACAGGCCAGCATCCGAATCTGCCAAGGACACGCTACTGCA 650
    |||
454 MetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluTh 470
    |||
651 CAAGATAGCCCGAAGTGGGGAAGGGGCGACCTCGACCA..... 689
    |||
470 rIleLeuLysAsnAsnIleSerSerGlyHisValProHisGly 484
    |||
690 .....CAGCGAAACATATCACAGCAACACCTCTCTCAGGGC 725

```


OM of: US-09-316-048-27 to: Issued_Patents_NA:* out_format : pfs
Date: Apr 5, 2001 8:34 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09316048/runat_01042001_170039_7753/app_query.fasta_1.1293
-DB=Issued_Patents_NA -Qfmt=fastap -SUFFIX=p2n.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blsum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09316048_@CGN1_1_106 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-316-048-27
Query length: 40
Database: Issued_Patents_NA:*
Database sequences: 280836
Database length: 80580151
Search time (sec): 112.270000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-651-818A-1	-	109.50	168.35	0.0799	2160	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-718-661-1	-	101.50	153.48	0.5375	2790	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-991-300-1	-	95.50	151.71	0.6744	1107	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-700-651-1	-	94.50	137.19	4.35	5163	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-928-361B-4	-	94.50	137.19	4.35	5163	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-700-651-2	-	94.50	136.94	4.49	5318	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-928-361B-3	-	94.50	136.94	4.49	5318	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-155-888-1	-	94.50	136.58	4.70	5552	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-13	+	94.00	128.02	14.08	13987	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-7	+	94.00	118.32	48.86	44377	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-198-1	+	94.00	118.32	48.86	44377	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-860-635A-18	-	92.00	140.20	2.95	2249	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-333A-1	-	92.00	136.51	4.74	3489	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-770-379-20	+	92.00	117.83	51.99	32207	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-860-635A-20	-	91.50	134.73	5.96	3923	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-928-361B-2	-	90.50	130.28	10.53	5511	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-928-361B-1	-	90.50	127.88	14.33	7334	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-205-506A-1	-	88.50	139.16	3.37	1312	!
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-02389-1	-	88.50	139.16	3.37	1312	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-146-930-1	+	88.50	125.68	19.01	6530	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-458-240-1	+	88.50	125.68	19.01	6530	!
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03993-1	+	88.50	125.68	19.01	6530	!
/cgn2_6/ptodata/2/ina/backfile1.seq:5168049-2	+	87.00	138.03	3.90	1131	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-609-716-40	+	83.00	145.23	1.55	225	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-475-411A-40	+	83.00	145.23	1.55	225	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-26	-	82.00	140.82	2.73	315	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-18	-	82.00	139.77	3.12	357	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-20	-	82.00	139.49	3.23	369	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-22	-	82.00	139.49	3.23	369	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-24	-	82.00	139.42	3.26	372	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-28	-	82.00	139.42	3.26	372	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-30	-	82.00	139.42	3.26	372	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-14	-	82.00	139.09	3.40	387	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-16	-	82.00	139.09	3.40	387	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-124-671-35	-	82.00	137.93	3.95	444	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-961-083-159	-	82.00	125.62	19.17	1924	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-864-038A-1	-	82.00	124.44	22.30	2214	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-864-038A-2	-	82.00	121.00	34.63	3331	!
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-864-038A-4	-	82.00	121.00	34.63	3331	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-1	+	81.50	98.67	607.65	43280	!

/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-433-954-3	-	80.50	126.56	16.99	1295
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-174-745A-3	-	80.50	126.56	16.99	1295
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-195-947-3	-	80.50	126.56	16.99	1295
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-433-885-3	-	80.50	126.56	16.99	1295
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-433-908B-3	-	80.50	126.56	16.99	1295

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-651-818A-1

seq_documentation_block:

Sequence 1, Application US/08651818A
Patent No. 5948889
GENERAL INFORMATION:
APPLICANT: de Boer, Piet A.J.
APPLICANT: Hale, Cynthia A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING
TITLE OF INVENTION: ANTIMICROBIALS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,818A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02249
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2160 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-651-818A-1

alignment_scores: Quality: 109.50 Length: 56
Ratio: 3.318 Gaps: 6
Percent Similarity: 58.929 Percent Identity: 50.000

alignment_block:

US-09-316-048-27 x US-08-651-818A-1/rev ..
Align seg 1/1 to reverse of: US-08-651-818A-1 from: 1 to: 2160

3 CysAlaAlaThrCysThr.....AlaGlyAlaGlyCysGly..... 14
||||:||||| ||||| ||||| ||||| |||||
1189 TGGCGTGTGATGACACAGCGCTGCGCGCTGCGCGACCTGTGGCGAAC 1140
15GlyCysCysGlyCysGly.....CysT 22
||||| ||||| ||||| |||||
1139 TGGCTGCTGCAACGCGCTGTTCAGGCTGCGCTGATAGGCACGCTGTGCA 1090
22 hr..CysAlaGlyAlaGly.....GlyGlyThrCysThrAlaGly 34
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1089 CCGGCTGCGCTGATGCGAGCATGTTCGCGGCTGACCTGCGCTTCAGGC 1040
35 ThrGlyCysGlyAlaGly 40

; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1

alignment_scores:

Quality:	94.50	Length:	31
Ratio:	4.109	Gaps:	1
Percent Similarity:	74.194	Percent Identity:	45.161

alignment_block:

US-09-316-048-27 x US-08-700-651-1/rev ..

Align seg 1/1 to reverse of: US-08-700-651-1 from: 1 to: 5163

7 CystHrAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcy 23
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
865 TGTAGTGTGTTGTTAGTGTGTTGTTGTTGTTGTTGTTGTTGTTG 816
23 sAlaGlyAlaGlyGlyGlyThrcCystHrAlaGlyThrcGlyCys 37
|::||| |::||| |::||| |::||| |::||| |::||| |::|||
815 TTGTGCT.....TGTAGTTGTGCTGTTGTTGTTGTTGTTGTTG 788

seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-928-361B-4

seq_documentation_block:

; Sequence 4, Application US/08928361B
; Patent No. 6071518

; GENERAL INFORMATION:

; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5163 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-928-361B-4

alignment_scores:

Quality:	94.50	Length:	31
Ratio:	4.109	Gaps:	1
Percent Similarity:	74.194	Percent Identity:	45.161

alignment_block:

US-09-316-048-27 x US-08-928-361B-4/rev ..

Align seg 1/1 to reverse of: US-08-928-361B-4 from: 1 to: 5163

7 CystHrAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcy 23
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
864 TGTAGTGTGTTGTTAGTGTGTTGTTGTTGTTGTTGTTGTTGTTG 815
23 sAlaGlyAlaGlyGlyGlyThrcCystHrAlaGlyThrcGlyCys 37
|::||| |::||| |::||| |::||| |::||| |::||| |::|||
814 TTGTGCT.....TGTAGTTGTGCTGTTGTTGTTGTTGTTGTTG 787

seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-700-651-2

seq_documentation_block:

; Sequence 2, Application US/08700651B
; Patent No. 6015882

; GENERAL INFORMATION:

; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

alignment_scores:

Quality:	94.50	Length:	31
Ratio:	4.109	Gaps:	1
Percent Similarity:	74.194	Percent Identity:	45.161

alignment_block:

US-09-316-048-27 x US-08-700-651-2/rev ..

Align seg 1/1 to reverse of: US-08-700-651-2 from: 1 to: 5318

7 CystHrAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcy 23
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
865 TGTAGTGTGTTGTTAGTGTGTTGTTGTTGTTGTTGTTGTTGTTG 816
23 sAlaGlyAlaGlyGlyGlyThrcCystHrAlaGlyThrcGlyCys 37
|::||| |::||| |::||| |::||| |::||| |::||| |::|||
815 TTGTGCT.....TGTAGTTGTGCTGTTGTTGTTGTTGTTGTTG 788

seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-928-361B-3

seq_documentation_block:

; Sequence 3, Application US/08928361B
; Patent No. 6071518

; GENERAL INFORMATION:

; APPLICANT: Petersen, Carolyn

```

; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-928-361B-3

alignment_scores:
      Quality: 94.50      Length: 31
      Ratio: 4.109      Gaps: 1
      Percent Similarity: 74.194      Percent Identity: 45.161

alignment_block:
US-09-316-048-27 x US-08-928-361B-3/rev ..

Align seg 1/1 to reverse of: US-08-928-361B-3 from: 1 to: 5318

      7 CystHrAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcY 23
      |||:::||||:::||||| ||||| ||||| |||||:::|
      864 TGTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 815

      23 sAlaGlyAlaGlyGlyGlyThrcYsThrAlaGlyThrcGlyCys 37
      |::||| |||:::|||||:::|
      814 TTGTGCT.....TGTAGTTGTTGCTTGTAGTTGT 787

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-155-888-1

seq_documentation_block:
; Sequence 1, Application US/08155888
; Patent No. 6066623
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Hedstrom, Richard C.
; APPLICANT: Sedegah, Martha
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE PROTECTIVE
; TITLE OF INVENTION: AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
; TITLE OF INVENTION: DELIVERING POLYNUCLEOTIDE VACCINES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: Naval Medical Res. & Dev. Cmd.
; STREET: Bldg. 1, T-12 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/155,888
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: N.C. 75,851
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 295-1022
; TELEFAX: (202) 295-6759
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5552 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..755
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /label= CMV-IE
; OTHER INFORMATION: /note= "this feature acts as a promoter for any
; OTHER INFORMATION: downstream DNA sequence."
; OTHER INFORMATION: /citation= ([2])
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 933..2367
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 933
; OTHER INFORMATION: /function= "protein protective against malaria"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /number= 1
; OTHER INFORMATION: /label= IL2-CSP
; OTHER INFORMATION: /citation= ([1])
; OTHER INFORMATION:
; APPLICANT: Sedegah, Martha
; APPLICANT: Hedstrom, Richard C.
; APPLICANT: Hoffman, Stephen L.
; TITLE: Vaccination with Plasmodium yoelii CS protein
; TITLE: plasmid DNA protects against malaria
; JOURNAL: Science
; PUBLICATION INFORMATION:
; AUTHORS: Cullen, Bryan R.
; TITLE: TRANS-ACTIVATION OF HUMAN IMMUNODEFICIENCY
; TITLE: VIRUS OCCURS VIA A BIMODAL MECHANISM
; JOURNAL: CELL
; VOLUME: 46
; PAGES: 973-982
; DATE: 26 SEP-1986
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 4732
; US-08-155-888-1

alignment_scores:
      Quality: 94.50      Length: 39
      Ratio: 3.635      Gaps: 2
```



```

;
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

```

```

alignment_scores:
    Quality: 92.00      Length: 37
    Ratio: 3.538      Gaps: 0
    Percent Similarity: 70.270      Percent Identity: 37.838

```

alignment_block:

US-09-316-048-27 x US-08-770-379-20 ..

Align seg 1/1 to: US-08-770-379-20 from: 1 to: 32207

```

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysCys 17
  :::::::::::::::::::::|||||:::|||||:::|||||
20429 TCCTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 20478

```

```

17 sGlyCysGlyCysThrCysAlaGlyAlaGlyGlyGlyThrCysThrAlaG 34
  1 ||||| ||:::|||||::: ||:::|||||::: ||:::
20479 CTGTTGTGGCTCTGCTGCTGTTGTGGCTCTGCTGAGGGGCTCTGCTG 20528

```

```

34 1yThrglyCys 37
  ||::: |||
20529 GCTCCTGCTGT 20539

```

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-860-635A-20

seq_documentation_block:

```

; Sequence 20, Application US/08860635A
; Patent No. 6143878

```

GENERAL INFORMATION:

```

; APPLICANT: Koopman, Peter
; APPLICANT: Goodfellow, Peter
; TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
; TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11530

```

COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,635A
; FILING DATE: 29-MAY-1997
; CLASSIFICATION: 514

```

PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: AU PM9714
; FILING DATE: 29-NOV-1994
; APPLICATION NUMBER: AU PM9835
; FILING DATE: 05-DEC-1994
; APPLICATION NUMBER: PCT/AU95/00799
; FILING DATE: 29-NOV-1995

```

ATTORNEY/AGENT INFORMATION:

```

; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX:

```

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

```

;
;   LENGTH: 3923 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-860-635A-20

```

```

alignment_scores:
    Quality: 91.50      Length: 44
    Ratio: 3.519      Gaps: 2
    Percent Similarity: 59.091      Percent Identity: 40.909

```

alignment_block:

US-09-316-048-27 x US-08-860-635A-20/rev ..

Align seg 1/1 to reverse of: US-08-860-635A-20 from: 1 to: 3923

```

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAla..... 11
  ::::::::::::::::::::: ||| :::|||||
1504 AGCGTGTGCGCCTGTGCTGCTGCGGGGCTGCCGCCGCTGCTG 1455

```

```

12 .....GlyCysGlyGlyCysCysGly.....CysGlyC 21
  ||||| ||||| ||||| ||||| |||||
1454 CGCCGCCCTGCGGCTGCGGGGCGCCCTGCGGGGCGCGGCGCTG 1405

```

```

21 ysthrCysAlaGlyAlaGlyGlyGlyThrCys 31
  ||::: ||::: ||||| ||::: |||
1404 GCTGCTGCGGGGCTGCGGGGCGCGGCGCTGCTGC 1373

```



```

KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
REFERENCE     1 (bases 1 to 295)
AUTHORS       Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
TITLE         The WashU-HMI Mouse EST Project
JOURNAL       Unpublished (1996)
COMMENT       Contact: Marra M/Mouse EST Project
              WashU-HMI Mouse EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LNL ; contact the
              IMAGE Consortium (info@image.lnl.gov) for further information.
              MGI:374089
              Seq primer: -28m13 rev1 ET from Amersham
              High quality sequence stop: 291.
FEATURES
  source
    1..295
      /organism="Mus musculus"
      /strain="C57BL/6"
      /db_xref="taxon:10090"
      /clone="IMAGE:608657"
      /clone_lib="Stratagene mouse skin (#937313)"
      /sex="females"
      /tissue_type="whole skin"
      /dev_stage="11 weeks old"
      /lab_host="SOLR (kanamycin resistant)"
      /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTT 3'"
BASE COUNT   31 a      70 c      132 g      61 t      1 others
ORIGIN
alignment_scores:
  Quality:   119.50      Length:   48
  Ratio:     3.621       Gaps:     2
  Percent Similarity: 68.750   Percent Identity: 43.750
alignment_block:
  US-09-316-048-27 x AA162430 ..
Align seg 1/1 to: AA162430 from: 1 to: 295
      1 ThrAlAcysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysCy 17
        ::::::::::: || :::::::::::|||||||
151 TCTAGCTGCGGAGGAGGCTGC..TGCGGGTCTGCTGCGGGGATGCCG 197
      17 sGlyCysGlyCysThrCysAlaGlyAlaGly..... 27
        |||||
198 AGGCTGCGGCTCTAGCTGCTGCGGATCTGCTGCGAGAGAGCTGCGGG 247
      28 .....GlyGlyThrCysThrAlaGlyThrGlyCysGlyAlaGly 40
        ||||| || :::::::::::||||| ::|||
248 GCTGTGAGAGCGGCTGCTGTGCGCGCTCCGGATGCTGCGGTGT 291
seq_name: gb_est12:AA798861
seq_documentation_block:
LOCUS      AA798861      355 bp      mRNA      EST      10-FEB-1998
DEFINITION vv94a10.r1 Stratigene mouse skin (#937313) Mus musculus cDNA clone

```

```

IMAGE:1230042 5', mRNA sequence.
ACCESSION      AA798861
VERSION        AA798861.1  GI:2861816
KEYWORDS
SOURCE
ORGANISM       Mus musculus
                house mouse.
REFERENCE
AUTHORS        Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
                Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
                Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
                Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
                Waterston,R.
TITLE          The WashU-HHMI Mouse EST Project
JOURNAL        Unpublished (1996)
COMMENT        Contact: Marra M/Mouse EST Project
                WashU-HHMI Mouse EST Project
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: mouseest@watson.wustl.edu
                This clone is available royalty-free through LNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                MGI:655634
FEATURES
source         Seq primer: -28m13 rev1 ET from Amersham
                High quality sequence stop: 249.
                Location/Qualifiers
                1..355
                /organism="Mus musculus"
                /strain="C57BL/6"
                /db_xref="taxon:10090"
                /clone="IMAGE:1230042"
                /clone_1lb="Stratagene mouse skin (#937313)"
                /sex="females"
                /tissue_type="whole skin"
                /dev_stage="11 weeks old"
                /lab_host="SOLR (kanamycin resistant)"
                /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
                ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
                dT. Whole skin from 11 week old C57BL/6 female mice.
                Average insert size: 1.0 kb; Un1-ZAP XR Vector: ~5'
                adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor
                sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT     19 a          161 g          101 t
ORIGIN
alignment_scores:
                Quality: 117.00          Length: 39
                Ratio: 4.034            Gaps: 3
Percent Similarity: 74.359      Percent Identity: 56.410
alignment_block:
US-09-316-048-27 x AA798861 ..
Align seg 1/1 to: AA798861 from: 1 to: 355
        3 CysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysGlyCys 19
          |||:||||: ||| |||:||||| ||| ||| ||| ||| |||
        211 TGCTGTGGCTGCTGT...GCTGTGTGGCTGTGTGGCTGTGTGGCTG 254
        19 s...GlyCysThrCysAlaGlyAlaGlyGlyGlyThrCysThrAlaGlyT 35
          | ||||| |||:||||:|||| ||| ||| ||| ::|||
        255 TGTGTGCTGTGTGGCTGTGGCTGTGTGGCTGTGGCTGTGTGGCTGTGCC. 303
        35 hrGlyCysGlyAlaGly 40
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        304 ..GTTGTGTGGCTGCGGT 318
seq_name: gb_est21:AI509887

```

seq_documentation_block:	293 bp	mrna	EST	15-MAR-2000
LOCUS	AI509887			
DEFINITION	ms32d09.y1 Stratagene mouse skin (#937313)		Mus musculus	cDNA clone
	IMAGE:608657 5', mRNA sequence.			
ACCESSION	AI509887			
VERSION	AI509887.1	GI:4408792		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.			
TITLE	The WashU-NCI Mouse EST Project 1999			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Marra M/WashU-NCI Mouse EST Project 1999			

correct orientation)
MGI:374089
Seq primer: -40RP from Gibco
High quality sequence stop: 272
POLYA=No.

FEATURES	Location/Qualifiers
source	1. .293

```

/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:608657"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcorRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GATTCGGCACGAG 3'-3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"

```

alignment_scores:		
Quality:	116.50	
Ratio:	3.758	
Percent Similarity:	68.889	
	Length:	45
	Gaps:	2
	Percent Identity:	44.444

```
alignment_block:
US-09-316-048-27 x AI509887
```

Align seg 1/1 to: A1509887 from: 1 to: 293

```

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysCys 17
   :::::||||:::  ||  :::::|||||
150 TCTAGCTGCGGAGGAGGCTGC... TCGGGTCTGCGCTCGGGGCGATGCGG 196
   |||
17 sGlyCysGlyCysThrCysAlaGlyAlaGly..... 27
   |||||
197 AGCGTCGCGGCTCTAGCTGCTCGGATCTGCGTCGCGAGAGCGCTCGGGG 246
   |||||

```

28glyglythrCystHrAlaglyThrGlyCys 37
 ||||| ||| :::::|||||
 247 GCTGTGGAGCGCGTTGCTGTGGCGGCTCCGGATGC 281

seq_name: gb_est11:AA727427

seq_documentation_block:			
LOCUS	AA727427	242 bp	EST
DEFINITION	vu96a04.r1 Stratagene mouse skin (#937313) Mus musculus cdna clone		
IMAGE:	1209966 5', mRNA sequence.		
ACCESSION	AA727427		
VERSION	AA727427.1	GI:2745134	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		

REFERENCE
AUTHORS
1 (bases 1 to 242)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE	The Washu-HIMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseeest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:646310

Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 200.

FEATURES	Location/Qualifiers
source	1. . 242

```

/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1209966"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dr. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GAATTCGCACGAG 3' ~3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

```

BASE COUNT	11 a	43 c	112 g	76 t
ORIGIN				

```

alignment_scores:
  Quality: 113.50
  Ratio: 3.783
  Percent Similarity: 71.429
  Length: 42
  Gaps: 2
  Percent Identity: 50.000

```

```
alignment_block:
US-09-316-048-27 x AA727427
```

Align seg 1/1 to: AA727427 from: 1 to: 242

```

3  CysAlaAlaIaThrCysThrAlaGlyAlaGlyCysGlyCys..... 16
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
99 TGTGTGGCTGTGGCTGCGGTGGCTGTGGCTGTGGCTGTGGCTGTGG 148
17 ....CysGlyCysGlyCysThrCysAlaGlyAlaGlyGlyGlyThrCysT 32

```

149 TGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTG 198
32 hrAlaGlyThrGlyCysGlyAlaGly 40
199 GCTGTGCT..GGTTGTGGCTGTGCT 221

seq_name: gb_est2:AA142537

seq_documentation_block:

LOCUS AA142537 387 bp mRNA EST 12-FEB-1997
DEFINITION ms07a04.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:606222 5', mRNA sequence.

ACCESSION AA142537

VERSION AA142537.1 GI:1711980

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through ILNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:371654

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 349.

Location/Qualifiers

FEATURES

source

1. 387

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:606222"

/clone_lib="Stratagene mouse skin (#937313)"

/sex="females"

/tissue_type="whole skin"

/dev_stage="11 weeks old"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI

; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo

dt. Whole skin from 11 week old C57BL/6 female mice.

Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'

adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 33 a 88 c 156 g 110 t

ORIGIN

alignment_scores:

Quality: 113.50

Ratio: 3.547

Percent Similarity: 50.794

Length: 63

Gaps: 3

Percent Identity: 36.508

alignment_block:

US-09-316-048-27 x AA142537 ..

Align seg 1/1 to: AA142537 from: 1 to: 387

3 CysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyGlyCysCysGlyCY 19

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

92 TGTGGCTGCGGTGGCTGTGCTGTGGCTGTGGCTGTGGCGGCTGTGCTTG 141
19 s.....GlyCysThrCysAlaGlyAlaGly.....GlyGlyThrC 31
142 TGGCTGTGGCTGTGGCTGTGGCGGCTGTGGCTGTGCTGTGGCTGCT 191

31 ys..... 31

192 GTGGCTGTGCTGTGGCTGTGCAGGCTTACAGTACTGTGCTGTGCGCGCGC 241

32ThrAlaGlyThrGlyCysGlyAlaGly 40

242 TGGCTGTGCGGCTTCCTGTGGCTGTGGCTGTGGCTGTGGC 280

seq_name: gb_gss14:AQ977896

seq_documentation_block:

LOCUS AQ977896 518 bp DNA GSS 29-JAN-2000
DEFINITION RPCI-23-336N5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-336N5,
DNA sequence.

ACCESSION AQ977896 GI:6810116

VERSION AQ977896

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 518)

Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinet
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Ressea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 336 row: N column: 5

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

FEATURES

source

1. 518

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-336N5"

/clone_lib="RPCI-23"

/sex="female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methyase. Size

selected DNA was cloned into the pBACe3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 148 a 190 c 115 g 65 t

ORIGIN

alignment_scores:

Quality: 113.50

Ratio: 3.439

Percent Similarity: 49.254

Length: 67

Gaps: 4

Percent Identity: 35.821


```
/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor
sequence: 5' CTCGAGTTT TTT TTT TTT TTT TTT TTT 3'"
BASE COUNT      20 a      48 c      85 g      68 t
ORIGIN

alignment_scores:
    Quality: 111.00      Length: 42
    Ratio: 3.700      Gaps: 4
    Percent Similarity: 71.429      Percent Identity: 54.762

alignment_block:
US-09-316-048-27 x AA733514 ..

Align seg 1/1 to: AA733514 from: 1 to: 221

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysCys 17
   :::::| | | | | | | | | | | | | | | | | | | | | |
87 AGCAGCTGC..ACCACCTGCAGATACCTACCGCTGCTGTGCTGCTG 133

17 sgly...CysGlyCysThrCysAlaGlyAlaGlyGly...GlyThrCysT 32
   ||| | | | | | | | | | | | | | | | | | | | | |
134 TGGCTGCTGTGGCTGTGGCTGCTGTGCGGTGGCTGTGCTGTGCTGTG 183

32 hrAlaGlyThrGlyCysGlyAlaGly 40
   ::| | | | | | | | | | | | | | | | | | | | | |
184 GCTGTGCT...GGCTGTGCTGTGCTG 206

seq_name: gb_est8:AA530753

seq_documentation_block:
LOCUS      AA530753      521 bp      mRNA      EST      22-JUL-1997
DEFINITION vj44e07.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:931908 5', mRNA sequence.

ACCESSION      AA530753
VERSION        AA530753.1 GI:2273491
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 521)
AUTHORS        Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
               Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
               Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
               Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
               Waterston,R.
TITLE          The WashU-HMI Mouse EST Project
JOURNAL        Unpublished (1996)
COMMENT        Contact: Marra M/Mouse EST Project
               WashU-HMI Mouse EST Project
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: mouseest@watson.wustl.edu
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               MGI:536828
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 298.
Location/Qualifiers
1. 521
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:931908"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"

FEATURES
source
```

```
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor
sequence: 5' CTCGAGTTT TTT TTT TTT TTT TTT TTT 3'"
BASE COUNT      62 a      122 c      188 g      149 t
ORIGIN

alignment_scores:
    Quality: 111.00      Length: 42
    Ratio: 3.700      Gaps: 4
    Percent Similarity: 71.429      Percent Identity: 54.762

alignment_block:
US-09-316-048-27 x AA530753 ..

Align seg 1/1 to: AA530753 from: 1 to: 521

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysCys 17
   :::::| | | | | | | | | | | | | | | | | | | | | |
4 AGCAGCTGC..ACCACCTGCAGATACCTACCGCTGCTGTGCTGCTG 50

17 sgly...CysGlyCysThrCysAlaGlyAlaGlyGly...GlyThrCysT 32
   ||| | | | | | | | | | | | | | | | | | | | | |
51 TGGCTGCTGTGGCTGTGGCTGCTGTGCGGTGGCTGTGCTGTGCTGTG 100

32 hrAlaGlyThrGlyCysGlyAlaGly 40
   ::| | | | | | | | | | | | | | | | | | | | | |
101 GCTGTGCT...GGCTGTGCTGTGCTG 123

seq_name: gb_est29:AU061961

seq_documentation_block:
LOCUS      AU061961      319 bp      mRNA      EST      20-MAY-1999
DEFINITION AU061961 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLG781, mRNA sequence.

ACCESSION      AU061961
VERSION        AU061961.1 GI:4883065
KEYWORDS      EST.
SOURCE        Dictyostelium discoideum.
ORGANISM      Dictyostelium discoideum
               Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE      1 (bases 1 to 319)
AUTHORS        Yoshino,R., Morio,T. and Tanaka,Y.
TITLE          Developmental cDNA in Dictyostelium discoideum
JOURNAL        Unpublished (1997)
COMMENT        Contact: Hideko Urushihara
               Institute of Biological Sciences
               University of Tsukuba
               3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
               Email: d402nu@sakura.cc.tsukuba.ac.jp
               PROJECT = Dictyostelium discoideum cDNA project in Japan.
Location/Qualifiers
1. 319
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLG781"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"

FEATURES
source

BASE COUNT      188 a      102 c      6 g      23 t
ORIGIN

alignment_scores:
    Quality: 110.50      Length: 40
    Ratio: 3.810      Gaps: 1
    Percent Similarity: 72.500      Percent Identity: 47.500
```



```
alignment_scores:
  Quality: 109.50      Length: 33
  Ratio: 4.212        Gaps: 3
  Percent Similarity: 78.788      Percent Identity: 63.636

alignment_block:
  US-09-316-048-27 x AA498971 ..

Align seg 1/1 to: AA498971 from: 1 to: 399

10 GtAlaAGlYcYsGtYcYcYs...CysGtYcYs...GtYcYsThrCysAl 24
   |||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22 GGCTGCGGCTGTGCTGGCTGTGCGCTGTGCGGCTGTGCGCTGTGG 71
   :|||::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
24 agtYAlaAGlYgltYthrCysThrAlaAGlYthrGtYcYsGtYAlaAGlY 40
   :|||::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
72 CGGCTGTGGCTGTGCGGCTGTGCGCTGTGGC...GGCTGTGGCTGTGGC 117

seq_name: gb_gss16:AZ123987

seq_documentation_block:
LOCUS      AZ123987      507 bp      DNA      GSS      12-MAY-2000
DEFINITION  RPCI-23-22G19_TV RPCI-23 Mus musculus genomic clone RPCI-23-22G19,
            DNA sequence.
ACCESSION   AZ123987
VERSION     AZ123987.1 GI:7793201
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 507)
AUTHORS     Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret,
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-23
            Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-22G19_TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 22 row: G column: 19
            Seq primer: SP6
            Class: BAC ends.

FEATURES
  source
    1..507
      location/Qualifiers
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPCI-23-22G19"
        /clone_lib="RPCI-23"
        /sex="Female"
        /lab_host="DH10B"
        /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
            EcORI; Site_2: EcORI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcORI and EcORI Methyase. Size
            selected DNA was cloned into the pBACe3.6 vector at the
            EcORI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
```

```
alignment_scores:
  Quality: 109.50      Length: 71
  Ratio: 3.422        Gaps: 3
  Percent Similarity: 45.070      Percent Identity: 32.394

alignment_block:
  US-09-316-048-27 x AZ123987/rev ..

Align seg 1/1 to reverse of: AZ123987 from: 1 to: 507

3 CysAlaAlaThrCysThrAlaAGlYAlaAGlYcYsGtYcYcYs..... 16
   |||::: ||| ::| ::| ||| ||| ||| ||| ||| ||| |||
454 TGCTGTGGTGGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGGCTGGC 405
   :|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 .....CysGtYcYs...GtYcYsThrCysAlaAGlYAlaAGlYgltY. 28
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 TGCTGTGGCTGTGCTGTGCGGCTGTGCTGTGCTGTGCGCCTGTGCTGTGGCT 355
   .....
28 .....
354 GTTGTGGCTGTGCTGTGCTGTGCTGTGCGCCTACGCTAGTCTGCTGTGCCGC 305
29 .....GlyThrCysThrAlaAGlYthrGl 36
   |||::: ||| ::| ||| ||| ||| ||| ||| ||| ||| |||
304 CGCAGCTGTGCTGCCCTCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGGCTGTGG 255
36 YcYsGtYAlaAGlY 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 CTGTGCTGTGGC 242

seq_name: gb_gss21:AZ434703

seq_documentation_block:
LOCUS      AZ434703      485 bp      DNA      GSS      03-OCT-2000
DEFINITION  IM0221C24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0221C24 F, DNA sequence.
ACCESSION   AZ434703
VERSION     AZ434703.1 GI:10558716
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 485)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0221 row: C column: 24
            Seq primer: CGTTGTAAACGACGCCACGT
            Class: plasmid ends
            High quality sequence stop: 485.

FEATURES
  source
    1..485
      location/Qualifiers
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="UUGC1M0221C24"
        /clone_lib="Mouse 10kb plasmid UUGC1M library"
        /sex="Male"
```

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g147321149b/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 62 a 82 c 206 g 135 t
ORIGIN

alignment_scores:
Quality: 109.00 Length: 50
Ratio: 3.303 Gaps: 2
Percent Similarity: 66.000 Percent Identity: 38.000

alignment_block:
US-09-316-048-27 x AZ434703 ..

Align seg 1/1 to: AZ434703 from: 1 to: 485

1 ThrAlaCysAlaAlaThrcystThrAlaGlyAlaGlyCysGlyCysCy 17
:::|||||::: ||| :::::|||||::: ||
49 AGTGGCTGTAGCAGTGGCTGTGGCTGTAGCAGTGGCTGTGGCAGTGGCTG 98
17 sGlyCysGlyCys.....ThrcysAlaGlyAlaGlyGlyGlyThr. 30
||||| ||| ::|||::|||::|||
99 TGGCTGTGGCTGTGGCAGTGGCTGTAGCAGTGGCTGTGGCAGTGGTGTG 148
31CysthrAlaGlyThrGlyCysGlyAlaGly 40
|||:: |||::|||::|||::|||
149 GCAGTGGCTATAGCAGTGGCTGCAGCTATGGCAGTGGCTGTGGTGTGC 198

OM of: US-09-316-048-27 to: N_Geneseq_36:* out_format : pfs
Date: Apr 5, 2001 8:30 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09316048/runat_01042001_170039_7768/app_query.fasta_1.1293
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09316048_ecgn1_1_248 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-316-048-27
Query length: 40
Database: N_Geneseq_36:*
Database sequences: 480022
Database length: 187831343
Search time (sec): 885.780000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V00469 -		109.50	161.66	0.3240	2160 ! H
/SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:N90746 -		105.00	161.87	0.3154	900 ! D
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X39707 -		104.50	163.22	0.2652	691 ! G
/SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:Q26509 -		102.50	148.56	1.74	3000 ! B
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V18480 -		101.50	147.64	1.96	2790 ! E
/SID6/gcgdata/geneseq/geneseqn/NA1990.DAT:Q03665 -		100.00	146.37	2.30	2465 ! S
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:T75505 -		98.00	160.32	0.3846	292 ! P
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z49963 -		98.00	160.32	0.3846	292 ! W
/SID6/gcgdata/geneseq/geneseqn/NA1991.DAT:Q14320 -		97.50	141.06	4.55	3000 ! F
/SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:N90096 -		96.50	132.86	13.03	6978 ! S
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V38816 -		95.50	145.99	2.42	1107 ! D
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z93361 -		94.50	149.46	1.55	592 ! Se
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V20700 -		94.50	132.25	14.08	5163 ! C
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V20701 -		94.50	132.02	14.51	5318 ! O
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:T80415 +		94.00	123.58	42.79	13987 !
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:T78508 +		94.00	114.41	138.80	44377 !
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:T80414 +		94.00	114.41	138.80	44377 !
/SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:T75504 -		93.00	154.22	0.8410	245 ! P
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:V73805 +		92.00	113.96	147.06	32207 !
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V19941 +		92.00	102.43	645.49	137507 !
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:T30308 -		91.50	133.47	12.04	2514 ! M
/SID6/gcgdata/geneseq/geneseqn/NA1991.DAT:Q14319 -		91.50	132.07	14.41	3000 ! F
/SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:Q34566 -		91.50	132.07	14.41	3000 ! F
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:T30309 -		91.50	129.92	19.00	3934 ! H
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V50431 -		91.50	125.12	35.14	7193 ! S
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:Z42235 -		90.00	135.19	9.66	1526 ! H
/SID6/gcgdata/geneseq/geneseqn/NA1994.DAT:Q73383 -		88.50	134.15	11.04	1312 ! H
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:T44325 -		88.50	134.15	11.04	1312 ! H
/SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:Q51557 -		88.50	121.39	56.68	6530 ! D
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z22072 +		88.50	121.39	56.68	6530 ! D
/SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:N93636 -		88.00	135.43	9.37	1016 ! S
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z51983 +		87.50	131.05	16.43	1604 ! M
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:A02294 -		86.00	135.18	9.68	719 ! H
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z45837 -		86.00	124.32	38.97	2821 ! C
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X60096 +		86.00	110.00	244.33	17091 !
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X33466 +		85.50	136.87	7.79	529 ! YL
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V03553 -		85.50	109.97	245.40	15620 !
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:A31042 -		85.00	148.17	1.83	116 ! PL
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:A00746 -		84.00	139.26	5.73	295 ! HU

/SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:Q41087 -	84.00	121.27	57.59	2838
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:Z09483 -	84.00	120.54	63.26	3112
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z33362 -	83.50	122.53	48.97	2203
/SID6/gcgdata/geneseq/geneseqn/NA1990.DAT:Q04710 +	83.00	139.91	5.27	225 !
/SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:T43976 +	83.00	139.91	5.27	225 !

seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V00469

seq_documentation_block:

ID V00469 standard; DNA; 2160 BP.
XX
AC V00469;
XX
DT 21-MAY-1998 (first entry)
XX
DE E. coli zipa gene.
XX
KW zipa gene; antimicrobial; antibiotic; screening; ss.
XX
OS Escherichia coli.
XX
FH Key location/Qualifiers
FT CDS 717..1703
FT sig_peptide 717..782
FT mat_peptide 738..1700
FT /*tag= c
XX
PN W09744481-A1.
XX
PD 27-NOV-1997.
XX
PE 21-MAY-1997; 97WO-US08703.
PR 21-MAY-1996; 96US-0651818.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI De Boer PAJ, Hale CA;
XX
DR WPI; 1998-018532/02.
DR P-PSDB; W36984.
XX
PT Method for screening for antimicrobial compounds - using the zipa
PT gene of E. coli
XX
PS Claim 12; Fig 4; 55pp; English.

This nucleotide sequence comprises the previously unidentified zipa gene located at 52 minutes on the Escherichia coli chromosome. A lambda gt11 library made from E. coli chromosomal DNA was probed with HKF1-FtsZ to identify clones capable of expressing zipa. 7 Recombinant phage were identified, all of which contained 7-10 kb of E. coli DNA, including the complete 2160 bp zipa gene. The gene encodes zipa protein (see W36984) that interacts with the essential division protein FtsZ. The invention relates to screening compounds for antimicrobial activity, especially by using bacterial proteins in vitro to detect compounds that interfere with cell division. An expression vector has been constructed to allow for zipa protein to be functionally over-expressed in bacterial cells or other hosts. The zipa protein can be used in a cell-free assay to screen compounds for their antimicrobial activity, and may also be used in a cellular assay such as a yeast two-hybrid system. The antimicrobial compounds identified by the method are effective against bacteria, and may also have activity against fungi, mycoplasma and protozoa.

Sequence 2160 BP; 587 A; 546 C; 548 G; 479 T; 0 other;

alignment_scores:

Quality: 109.50 Length: 56

Ratio: 3.318 Gaps: 6
Percent Similarity: 58.929 Percent Identity: 50.000

alignment_block:

US-09-316-048-27 x V00469/rev ..

Align seg 1/1 to reverse of: V00469 from: 1 to: 2160

```
3 CysAlaAlaThrCysThr.....AlaGlyAlaGlyCysGly..... 14
||||:||||| ||||| ||||| |||||
1189 TGGCGTGTGAATGCACAGGCTGCGGCTGGCGCGACCTGTGGCGAAAC 1140
15 .....GlyCysCysGlyCysGly.....CysT 22
||||| ||||| ||||| |||||
1139 TGGCTGTGCAACGGCTGTTCAGGCTGCGGCTGATAGGACGGCTGTGCA 1090
22 hr...CysAlaGlyAlaGly.....GlyGlyThrCysThrAlaGly 34
|| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1089 CCGGCTGCGCTGATGCGGAGCATGTGCGGCGGTACCTGCGCTTCAGGC 1040
35 ThrGlyCysGlyAlaGly 40
||||| :|||
1039 ...GGCTGTGACCGGT 1025
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:N90746
```

seq_documentation_block:

ID N90746 standard; DNA; 900 BP.

AC N90746;

DT 06-JUN-1990 (first entry)

DE DNA encoding the Plasmodium berghei circumsporozoite protein.

XX Plasmodium berghei; malaria; circumsporozoite protein; live

KW recombinant vaccine; Salmonella; epitope.

XX Plasmodium berghei.

FT Key Location/Qualifiers

FT CDS 1..820

FT repeat_region /*tag= a

FT repeat_region 121..192

FT repeat_region /*tag= b

FT repeat_region /rpt_unit=GACCCACACCAACCAAAACCAAAAT

FT repeat_region 193..360

FT repeat_region /*tag= c

FT repeat_region /rpt_unit=GACCCACACCAACCAAAACCAAAAT

FT repeat_region 415..480

FT repeat_region /*tag= d

FT repeat_region /note="region encoding repeated dipeptide units"

PN WO8902924-A.

PD 06-APR-1989.

PF 30-SEP-1988; 88WO-US03376.

PR 02-OCT-1987; 87US-0104735.

PA (PRAX-) PRAXIS BIOLOGICS IN.

PI Brey RN, Majarian WR, Pillal S, Hockmeyer WT;

DR WPI; 1989-114399/15.

DR P-PSDB; P93560.

PT Live recombinant vaccine for malaria -

PT comprising attenuated entero-invasive bacterium contg. DNA

PT encoding epitope of malaria parasite

PS Fig 2; p. 2/17; 105pp; English.

XX Circumsporozoite protein (CS) is an epitope of the malarial parasite,
CC Plasmodium berghei. In the patent, the sequence is inserted into an
CC attenuated enteroinvasive bacterium where it is expressed. Such bacteria
CC can multiply in the host without causing disease or disorder and express
CC CS that will induce a protective immune response against malaria and
CC can be used in vaccines. Such vaccines can be multivalent.

SO Sequence 900 BP; 378 A; 236 C; 126 G; 160 T; 0 other;

alignment_scores:

Quality: 105.00 Length: 31

Ratio: 4.565 Gaps: 0

Percent Similarity: 74.194 Percent Identity: 54.839

alignment_block:

US-09-316-048-27 x N90746/rev ..

Align seg 1/1 to reverse of: N90746 from: 1 to: 900

```
10 GlyAlaGlyCysGlyGlyCysCysGlyCysGlyCysThrCysAlaGly 26
||||:||||| ||||| ||||| ||||| ||||| ||||| |||||
506 GGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 457
26 aGlyGlyGlyThrCysThrAlaGlyThrGlyCysGlyAlaGly 40
: |||| ||| :||| ||||| ||||| ||||| |||||
456 CTGTGGCTGTGCTGTGCTGTGCTGTGCGCGCGCTGTGCTGTGTGCA 414
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:X39707
```

seq_documentation_block:

ID X39707 standard; DNA; 691 BP.

AC X39707;

DT 02-JUL-1999 (first entry)

DE Gastric cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer; ss.

XX Homo sapiens.

PN WO9904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Oyata Y, Old LJ;

PI Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

DR WPI; 1999-132448/11.

PT New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

PS Claim 67; Page 518; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

XX
SQ Sequence 691 BP; 162 A; 247 C; 139 G; 135 T; 8 other;

alignment_scores:
Quality: 104.50 Length: 36
Ratio: 4.354 Gaps: 2
Percent Similarity: 66.667 Percent Identity: 55.556

alignment_block:
US-09-316-048-27 x X39707/rev ..

Align seg 1/1 to reverse of: X39707 from: 1 to: 691

7 CysThrAlaGlyAlaGlyCysGlyGlyCys.....CysGlyCys 19
111 :::::111 11111111 11111111
106 TGCTGGGAGGGGGTGGGGTGGAGCTGCTGGGCTGAGGCTGCGGCTG 57
19 sGlyCysThrCysAlaGlyAlaGlyGlyGlyThrCysThrAlaGlyThrG 36
1 111 111::: ::::1111111 111 11111 1
56 CTGTGCGGCTGCTGCTGGGGTGGTGAGGC...TGCTGCTGCTGTTGTG 10
36 1yCysGly 38
11111111
9 GTTGTGGC 2

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT:Q26509

seq_documentation_block:
ID Q26509 standard; DNA; 3000 BP.

XX Q26509;
XX
DT 08-JAN-1993 (first entry)
XX
DE Bordetella parapertussis prn gene.
XX
KW Whooping cough; P70 antigen; P95 precursor protein; vaccination; ss.
XX
OS Bordetella parapertussis.

XX Key Location/Qualifiers
FH CDS 145..2913
FT /*tag= a
FT /product= P95_precursor
FT mat_peptide 247..2073
FT /*tag= b
FT /product= P70
FT repeat_region 940..999
FT /*tag= c
FT repeat_unit 940..954
FT /*tag= d
FT /rpt_type= DIRECT, TANDEM
FT repeat_region 1867..1980
FT /*tag= e
FT /note= "encodes 9 x Pro-Gln-Pro repeats"
FT repeat_unit 1867..1875
FT /*tag= f
FT /rpt_type= DIRECT

XX
PN WO9211292-A.
XX
PD 09-JUL-1992.
XX
PF 23-DEC-1991; 91WO-GB02302.
XX
PR 21-DEC-1990; 90GB-0027901.
XX
PA (WELL) WELLCOME FOUND LTD.
XX
PI Charles IG;
XX
DR WPI; 1992-250033/30.
DR P-PSDB; R25578.
XX

PT Acellular vaccine for immunisation against whooping cough -
PT comprises protein uncontaminated by B. parapertussis components
PT and capable of binding antibodies which bind native P70 antigen
XX
PS Claim 2; Fig 1; 20pp; English.

CC The DNA encodes the P95 precursor of B.parapertussis which is
CC processed to give the P70 antigen. A cosmid library was constructed
CC by transforming E.coli HB101 with recombinant cosmids prepared
CC by partial digestion of B.parapertussis chromosomal DNA
CC with Sau3A and cloning of 40-50kb fragments into the BamHI site of
CC cosmid pHC79. The cosmids were screened with a 1.8kb ClaI fragment
CC from the prn gene of B.pertussis. The insert from one positive
CC colony, harbouring cosmid pBD811, was sequenced and found to
CC contain an open reading frame encoding a 922 amino acid protein
CC with calculated mol.wt. 95,177.

XX
SQ Sequence 3000 BP; 451 A; 1025 C; 1102 G; 422 T; 0 other;

alignment_scores:
Quality: 102.50 Length: 43
Ratio: 3.942 Gaps: 2
Percent Similarity: 60.465 Percent Identity: 53.488

alignment_block:
US-09-316-048-27 x Q26509/rev ..

Align seg 1/1 to reverse of: Q26509 from: 1 to: 3000

4 AlaAlaThrCysThrAlaGlyAla.....GlyCysGly 14
111::: 111 11111111 11111111
1985 GCCGGCGGTGGCGGCCGCCGCTTCCGGCTGCCTGTGTGGCGGCTGCCG 1936
14 yGlyCysCysGlyCysGlyCysThrCysAlaGlyAlaGlyGlyThrC 31
111111 11111111 111:::11111 11111111
1935 CGGCTGGCGGCGGCTGCGGC...GGCTGGCGGCGGCTGGCGGCTGGGAC 1889
31 ySThrAlaGlyThrGlyCysGlyAlaGly 40
111 11111111111111111111
1888 CGGGCTGGGACCGGCTGCGGCGGCGG 1860

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V18480

seq_documentation_block:
ID V18480 standard; cDNA to mRNA; 2790 BP.

XX V18480;
XX
DT 18-AUG-1998 (first entry)
XX
DE BOP1 cDNA.
XX
KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;
KW inducible; alzheimer's disease; nuclear transcription factor; apoptosis;
KW cell cycle; neuronal disorder; ss.
XX

```
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 542..2545
FT /*tag= a
FT /product= "BOP1 protein"
XX
PN W09813489-A1.
XX
PD 02-APR-1998.
XX
PF 22-SEP-1997; 97WO-EP05198.
XX
PR 23-SEP-1996; 96US-0718661.
XX
PA (CNRS ) CENT NAT RECH SCI.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Journot L, Spengler D;
XX
DR WPI; 1998-230701/20.
DR P-PSDB; W48760.
XX
PT New isolated tumour suppressor gene - useful for developing products
PT for use in diagnosis and treatment of tumour(s) or neuronal
PT disorder(s)
XX
PS Claim 1; Pages 72-76; 118pp; English.
XX
CC The present sequence represents the BOP1 cDNA isolated from the
CC mouse corticotroph pituitary tumour cell line AtT-20 cDNA library.
CC The protein encoded by the BOP1 cDNA displays a tumour suppressing
CC activity when it was constitutively and inducibly expressed in
CC tumour cells. The BOP1 cDNA and the protein it encodes are claimed
CC to be useful in the preparation of therapeutic compositions, useful
CC for treating, preventing or delaying the recurrence of a tumour or
CC neuronal disorders, e.g. genetic diseases or acquired degenerative
CC encephalopathies such as Alzheimer's disease. The BOP1 protein is
CC also claimed to be able to induce apoptosis resulting in inhibition
CC of tumour cell growth, to suppress tumour formation, to induce G1
CC arrest of the cell cycle and to act as nuclear transcription factor.
XX
SQ Sequence 2790 BP; 667 A; 783 C; 714 G; 626 T; 0 other;

alignment_scores:
Quality: 101.50 Length: 47
Ratio: 2.985 Gaps: 4
Percent Similarity: 72.340 Percent Identity: 42.553

alignment_block:
US-09-316-048-27 x V18480/rev ..

Align seg 1/1 to reverse of: V18480 from: 1 to: 2790

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCys.. 16
:::|||||: |||:::|||||:|||||:|||||:
2286 TCAGGCTGCTGCTGAGTTGCAAGTTCGCGCAGCAGCTGTGTAGCTGCAT 2237
17 .....Cys.....GlyCysGlyCysThrCysAlaGlyAlaG 27
||| |||:::||||| ||| |||:::|
2236 CTGGGGCTGGATCTGCAACTGAACACTGTGGCTGCATCTGT..GGCTGTG 2190

27 lylGlyGlyThrCysThrAlaGlyThrGlyCysGlyAlaGly 40
||:::|||| ||| :::|||||:|||||:|||||
2189 GTAGTGGC..TGTGGCACTGGCAGTGGCTGTGGCAGTGGC 2152

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT:Q03665
seq_documentation_block:
ID Q03665 standard; DNA: 2465 BP.
XX
AC Q03665;
```

```
XX
DT 07-AUG-1989 (first entry)
XX
DE Sequence homologous to Drosophila Per gene.
XX
KW Sex determination; ruminant embryos; ss.
XX
PN FR2635116-A.
XX
PD 09-FEB-1990.
XX
PF 08-AUG-1988; 88FR-0010706.
XX
PR 08-AUG-1988; 88FR-0010706.
XX
PA (GEOR/) GEORGES M.
XX
PI Georges M, Vassart G, Christophe D, Dumont J, Young M;
XX
DR WPI; 1990-093373/13.
XX
PT Determn. of sex of ruminant embryos-by hybridisation assay using specific
PT probes.
XX
PS Claim 17; fig 1; 37pp; French.
XX
CC This sequence is used as a probe and is contained in plasmid pSP64.2 SEI.
CC It is homologous to the sequence of a 2.5 kb fragment of the Per gene of
CC Drosophila. The motifs ACONGN and/or TCAGGC (N=T, U, G, A, or C) or their
CC complementary sequences are repeated in the probe. It is used in a
CC hybridisation assay to determine the sex of ruminant embryos. Either
CC (i) Southern- or (ii) Dot-blot techniques can be utilised and Y-chromo-
CC some shows up as intense signal either obscuring most of the hybridisa-
CC tion track in (i) or surrounding the dots in (ii). Significant results can
CC be obtd. using samples of less than 50 ng weight.
XX
SQ Sequence 2465 BP; 791 A; 739 C; 689 G; 246 T; 0 other;

alignment_scores:
Quality: 100.00 Length: 36
Ratio: 4.000 Gaps: 0
Percent Similarity: 69.444 Percent Identity: 44.444

alignment_block:
US-09-316-048-27 x Q03665/rev ..

Align seg 1/1 to reverse of: Q03665 from: 1 to: 2465

2 AlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysG 18
|||||||:|||||:|||||:|||||:
593 GCCTGTGCCTGTGCTGTCTGTGACTTGTGCTGTGCTGTCTGTGTA 544
18 yCysGlyCysThrCysAlaGlyAlaGlyGlyThrCysThrAlaGlyT 35
|||||||:|||||:|||||:|||||
543 CTGTGGCTGTTCCTGTTCTGTGCTGTGCTGTGCTGTGACCGTGGCT 494
35 hrglyCys 37
:::|||||
493 CTGGCTGT 486

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T75505
seq_documentation_block:
ID T75505 standard; cDNA: 292 BP.
XX
AC T75505;
XX
DT 23-MAR-1998 (first entry)
XX
DE P. americanus skin type AFP, clone SAFP8 cDNA.
XX
KW Antifreeze polypeptide; intracellular skin type AFP; winter flounder;
```

```

KW cold resistance; thermal hysteresis; antibacterial; ss.
XX Pseudopleuronectes americanus.
OS
FH Key Location/Qualifiers
FT CDS 36..200
FT /*tag= a
FT /product= saFP8
FT /note= "skin type antifreeze polypeptide"
XX
PN W09728260-A1.
XX
PD 07-AUG-1997.
XX
PF 30-JAN-1997; 97WO-CA00062.
XX
PR 31-JAN-1996; 96US-0010920.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Gong Z, Hew C;
XX
DR WPI; 1997-402614/37.
DR P-PSDB; W22875.
XX
PT Isolated fish skin anti-freeze polypeptide - useful to depress
PT freezing points of aqueous compositions and protect plant, fungal,
PT animal or bacterial cells from cold
XX
PS Claim 11; Figure 11; 104pp; English.
XX
CC This sequence encodes the saFP8 clone corresponding to a novel skin type
CC intracellular antifreeze polypeptide (AFP) which induces a concentration
CC dependent decrease in the freezing point of an aqueous solution. This
CC novel polypeptide does not contain a signal sequence and is thought to be
CC intracellular. AFP's can be used to make an aqueous composition resistant
CC to freezing by changing its thermal hysteresis such as a water or salt
CC solution, an intracellular compartment of a cell or a food stuff, e.g.
CC soft serve 'frozen' yoghurt or ice cream. AFP's can inhibit ice
CC recrystallisation during cold storage, improving the texture and
CC palatability of the food and has antibacterial properties. Such
CC polypeptides can also be expressed to provide cold resistance to cells,
CC e.g. plant, fungal animal or bacterial cells. The antibodies can be used
CC to identify and isolate AFP while its promoter can be used to direct
CC expression of a nucleic acid.
XX
SQ Sequence 292 BP; 66 A; 109 C; 76 G; 41 T; 0 other;
XX
alignment_scores:
Quality: 98.00 Length: 33
Ratio: 4.261 Gaps: 3
Percent Similarity: 69.697 Percent Identity: 60.606
XX
alignment_block:
US-09-316-048-27 x T75505/rev ..
XX
Align seg 1/1 to reverse of: T75505 from: 1 to: 292
XX
10 gAlAalaglyCysglyGlyCys...Cysgly...CysglyCysThrCysAl 24
117 GGCAGTGGCTGGCGGGCTGGCGGTGCCAGCTGCGGTGCTTCGCGG 88
137 GGCAGTGGCTGGCGGGCTGGCGGTGCCAGCTGCGGTGCTTCGCGG 88
24 aglyAlaGlyGlyGlyThrCysThraAlaGlyThrGlyCysgAlaGly 40
117 GGCAGTGGCTGGCGGGCTGGCGGTGCCAGCTGCGGTGCTTCGCGG 88
87 CGGCTTGCGGGCGGC.....GGCGGTGGCTGGCGGGCGGC 51
XX
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:Z49963
seq_documentation_block:
ID Z49963 standard; CDNA; 292 BP.
XX
Z49963;
AC
```

```

XX      25-APR-2000 (first entry)
DT
XX
DE      Winter flounder skin-type antifreeze protein-8 DNA.
XX
KW      Winter flounder skin-type antifreeze protein; wfsAFP-8; antibacterial;
KM      cold tolerance; ice recrystallisation; refrigerated food;
KX      antifreeze; ds.
XX
OS      Pseudopleuronectes americanus.
XX
PN      WO200000512-A2.
XX
PD      06-JAN-2000.
XX
PF      25-JUN-1999;    99WO-CA00601.
XX
PR      26-JUN-1998;     98US-0090794.
PR      07-AUG-1998;     98US-0095713.
PR      24-JUN-1999;     99US-0344529.
XX
PA      (HSCR-) HSC RES & DEV LP.
XX
PI      Hew CL;
XX
DR      WPI; 2000-170905/15.
XX
PT      Novel antifreeze polypeptides and polynucleotides used to make cells
PT      cold resistant and to improve the palatability of cold foods and
PT      liquids -
PS
XX      Disclosure; Fig 5; 61pp; English.
XX
CC      The present sequence encodes winter flounder skin type antifreeze
CC      protein (wfsAFP-8). This lacks a signal peptide and is indicative of its
CC      intracellular location. Polyclonal antisera generated using sculpin skin
CC      type AFP-2 is tested for its cross-reactivity against wfsAFPs using a
CC      competitive binding immunoassay. The AFPs are used to make aqueous
CC      compositions resistant to freezing, to improve the cold tolerance, for
CC      inhibiting ice recrystallisation, and for stabilising biological
CC      membranes. They are also used to improve the shelf life and food quality
CC      of many refrigerated foods and have antibacterial properties.
XX
SQ      Sequence 292 BP; 66 A; 109 C; 76 G; 41 T; 0 other;

alignment_scores:
Quality:          98.00           Length:          33
Ratio:            4.261           Gaps:          3
Percent Similarity: 69.697       Percent Identity: 60.606

alignment_block:
US-09-316-048-27 x Z49963/rev ..

Align seg 1/1 to reverse of: Z49963 from: 1 to: 292

10 GLYALAGLYCysGLyGLyCys...CysGLy..CysGLyCysThrCysAl 24
   |||:::|||||:||||| ||||| ||||| |||::
137 GCGAGTGGCTGCGGCGGCTGCGGTGCAGTCGCGTGCCTTGCGGC 88

24 aclyAalaglyglyThrcystThralaglyThrglycysglaagly 40
   :||| |||||||| | | | | | | | | | | | | | | | | | |
87 CGGCCTTGGCGGCGGC.....GGCGGTGGCTGCGGCGGCGGC 51

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT:Q14320

seq_documentation_block:
ID Q14320 standard; DNA; 3000 BP.
AC Q14320;
XX
XX      20-JAN-1992 (first entry)
DT
XX
```

```
DE Pertactin antigen P.70.
XX
XX Pertactin; Pichia; B. pertussis; B.bronchiseptica; ss.
XX
XX Bordetella parapertussis.
OS
XX
FH Key Location/Qualifiers
FT CDS 145..2910
FT /*tag= a
FT /label= pertactin
XX
XX WO9115571-A.
XX
XX 17-OCT-1991.
XX
XX 28-MAR-1991; 91WO-GB00487.
XX
XX 02-APR-1990; 90GB-0007416.
XX
XX (WELL ) WELLCOME FOUNDATION LTD.
XX
XX PI Clare JJ, Romanos MA;
XX
XX WPI; 1991-325214/44.
XX
XX P-PSDB; R14321.
XX
XX Pichia microorganism transformants - for production of
PT Bordetella pertactin antigens for whooping cough vaccines
XX
XX PS Disclosure; Fig 1C; 38pp; English.
XX
XX Pichia microorganisms are transformed for the expression of
CC pertactin antigens. DNA sequence used are represented in Q14319-20
CC encoding the B. bronchiseptica P.68 and B. parapertussis P.70 antigen
CC respectively or the B. pertussis P.69 encoding sequence described
CC by I.G. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80:3554-3448
CC (1989).
CC The 109 Ns represent nucleotides missing in the specification.
XX
XX Sequence 3000 BP; 427 A; 987 C; 1073 G; 404 T; 109 other;
SQ

alignment_scores:
Quality: 97.50 Length: 42
Ratio: 3.611 Gaps: 2
Percent Similarity: 64.286 Percent Identity: 52.381

alignment_block:
US-09-316-048-27 x Q14320/rev ..
Align seg 1/1 to reverse of: Q14320 from: 1 to: 3000

4 AlaAlaThrcysThrAlaGlyAlaGly.....CysGlyGlyCyscy 17
|||::: ||| |||||::: |||||
1985 GCCGGCGGTGCGGCGCGGCTTCGGCTGCTGTGTGGCGGCTGCGG 1936

17 sGlyCysGlyCysThr.....CysAlaGlyAlaGlyGlyGlyThrCysT 32
|||||::: |||::: |||::: |||
1935 CGGCTGCGGCMNNMNCGGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1886

32 hrAlaGlyThrGlyCysGlyAlaGly 40
||| |||||
1885 GCTGGGACCGGCTGCGGCGGCGG 1860

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:N90096
seq_documentation_block:
ID N90096 standard; cDNA; 6978 BP.
XX
XX N90096;
XX
XX 22-MAR-1991 (first entry)
XX
```

```
DE Sequence of plasmid PAY31 encoding wheat alpha-gliadin.
XX
XX Coeliac disease diagnosis; dough formation; food technology; ds.
XX
XX Wheat.
XX
XX US4826765-A.
XX
XX 02-MAY-1989.
XX
XX 24-OCT-1986; 86US-09222616.
XX
XX 24-OCT-1986; 86US-09222616.
XX
XX 24-OCT-1986; 86US-09222616.
XX
XX (UYHA-) UNIV OF HAWAII (USDA).
XX
XX PI Greene FC, Stiles JT, Neill JD, Anderson OD, Litts JC;
XX
XX WPI; 1989-150270/20.
XX
XX New wheat gluten protein gene plasmid - useful for genetic
PT transformation of yeasts to produce gliadin or glutenin
XX
XX PS Disclosure; Fig 4; 21pp; English.
XX
XX The preferred gene fragment in the patent of the invention encodes
CC alpha-gliadin or glutenin, and the preferred plasmid encoding alpha-
CC gliadin is PAY 31. The plasmid may be used for the prodn. of alpha-
CC gliadin, which may be used in the food technology industry to modify
CC ,prodn. processes of doughs and batters and to reduce unit costs.
CC Gliadin may also be used for diagnosis and treatment of illness
CC caused by wheat gluten proteins and for testing theories of dough
CC formation.
XX
XX Sequence 6978 BP; 2042 A; 1631 C; 1447 G; 1858 T; 0 other;
SQ

alignment_scores:
Quality: 96.50 Length: 43
Ratio: 3.860 Gaps: 2
Percent Similarity: 58.140 Percent Identity: 44.186

alignment_block:
US-09-316-048-27 x N90096/rev ..
Align seg 1/1 to reverse of: N90096 from: 1 to: 6978

3 CysAlaAlaThrcysThrAlaGlyAlaGlyCysGly..... 15
||||::: |||::: ||||| |||
3080 TGCTGCTGCTGCTGCTGCGAATGCTGTGTGTCGGAATACTG 3031

16 CysCysGlyCysGlyCysThrCysAlaGlyAlaGlyGlyThrCysT 32
||||| ||||| |||::: |||
3030 TTGTTGCGGTGTGATGCTTGTGTGTCGAAATGTTGTGCTGTG 2981

32 hrAlaGlyThr...GlyCysGlyAlaGly 40
|||::: ||||| |||
2980 AATATGTAAGTTCGCGCTCGGAATGCT 2952

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:V38816
seq_documentation_block:
ID V38816 standard; DNA; 1107 BP.
XX
XX V38816;
XX
XX 09-OCT-1998 (first entry)
XX
XX DNA encoding a durum wheat glutenin protein.
DE
XX
XX Glutenin gene; durum wheat; low-molecular-weight;
KW transgenic durum wheat; ss.
XX
```


OS Triticum durum.
XX
PN FR2757538-A1.
XX
PD 26-JUN-1998.
XX
PF 18-DEC-1997; 97FER-0016059.
XX
PR 19-DEC-1996; 96IT-MI02663.
XX
PA (ITU-Y-) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
XX
PI Cardelli LE, D'Ovidio R, Marchitelli C, Porceddu E;
XX
DR WPI; 1998-365055/32.
DR P-PSDB; W62647.
XX
PT Durum wheat glutenin gene - coding for glutenin protein of low
PT molecular weight
XX
PS Claim 1; Page 13; 18pp; French.
XX
CC The present sequence represents a glutenin gene, and is isolated from
CC the genomic DNA of Triticum durum L. The gene codes for a
CC low-molecular-weight glutenin protein and can be used to produce
CC transgenic durum wheat plants with "better quality characteristics"
CC [no details given].
XX
SQ Sequence 1107 BP; 368 A; 377 C; 152 G; 210 T; 0 other;

alignment_scores:
Quality: 95.50 Length: 48
Ratio: 3.673 Gaps: 3
Percent Similarity: 54.167 Percent Identity: 41.667

alignment_block:
US-09-316-048-27 x V38816/rev ..

Align seg 1/1 to reverse of: V38816 from: 1 to: 1107

1 ThrAlaCysAlaAlaThr.....CysThrAlaGlyAlaGlyCy 13
||||:||||:||||: ||| :||| |||||
572 ACTGGCTGTGTGTCGAAATGTTGTTGCGGTGCGGAATTGTTG 523

13 S.....GlyGlyCysCysGly...CysGlyCysThrC 23
| ||||| ||||| ||||| |||||
522 TTGTTGTGTGTCGAAATGTTGTTGCGGTAGTACTGTTGTTGTT 473

23 ysAlaGlyAlaGlyGlyGlyThrCysThrAlaGlyThrGlyCys 37
||::: |||||::: ||| ||| |||||
472 GCTGCGAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 429

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:Z93361
seq_documentation_block:
ID Z93361 standard; DNA; 592 BP.
XX
AC Z93361;
XX
DT 16-AUG-2000 (first entry)
XX
DE Sequence encoding F-box protein FBP-11.
XX
KW F-box protein; FBP; diagnosis; treatment; screening; agonist;
KW antagonist; proliferative disorder; differentiative disorder;
KW breast cancer; prostate cancer; ovarian cancer; cancer;
KW small cell lung carcinoma; immune disorder; cardiovascular disorder;
KW inflammatory disorder; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

FT CDS 2..592
FT /*tag= a
FT /product= FBP-11
XX
PN WO200012679-A1.
XX
PD 09-MAR-2000.
XX
PF 27-AUG-1999; 99WO-US19560.
XX
PR 28-AUG-1998; 98US-0098355.
PR 03-FEB-1999; 99US-0118568.
PR 15-MAR-1999; 99US-0124449.
XX
PA (UYNY) UNIV NEW YORK STATE.
XX
PI Chaur DS, Pagano M, Latres E;
XX
DR WPI; 2000-256635/22.
DR P-PSDB; Y83079.
XX
XX
PT Novel nucleic acid for screening compounds useful for treating
PT proliferative and differentiative disorders such as cancer and immune
PT disorders comprises sequences encoding ubiquitin ligases -
XX
PS Claim 4; Figure 14b; 245pp; English.
XX
CC Nucleic acids encoding substrate-targeting subunits of ubiquitin
CC ligases with F-box motifs (F-box proteins) are useful for diagnosis
CC of proliferative and differentiated related disorders by measuring
CC FBP gene expression. Cells expressing such proteins or
CC their fragments are useful for screening compounds. The compounds
CC are agonists or antagonists, which are useful for treating a
CC proliferative or differentiative disorder in a mammal such as
CC breast, ovarian and prostate cancer and small cell lung carcinoma
CC and also major opportunistic infections, immune disorders,
CC cardiovascular diseases and inflammatory disorders. FBP protein,
CC analogs, derivatives and their subsequences, anti-FBP antibodies
CC are also useful in diagnosis of the disorders.
XX
SQ Sequence 592 BP; 160 A; 171 C; 148 G; 113 T; 0 other;

alignment_scores:
Quality: 94.50 Length: 34
Ratio: 3.938 Gaps: 3
Percent Similarity: 70.588 Percent Identity: 55.882

alignment_block:
US-09-316-048-27 x Z93361/rev ..

Align seg 1/1 to reverse of: Z93361 from: 1 to: 592

10 GLyAlaGlyCysGlyGlyCysCysGlyCysGlyCysThrCysAlaGlyAl 26
|||::||| ||||| ||||| ||| ||| |||::|||::
141 GTGGCGCGCGCGGAGGCTGCTGC...TGC...TGCTGCTGCTGCGCGG 98

26 aGlyGlyGlyThrCysThrAlaGlyThr.....GlyCysGlyAlaG 40
:||||||| ||| ::||| ||||| |||||::|
97 CGCGGAGGCTGCTGTTGGGCGGCTGCTGTTGGGCGGCTGCGCGGCG 48

40 ly 40
||
47 GC 46

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V20700
seq_documentation_block:
ID V20700 standard; DNA; 5163 BP.
XX
AC V20700;
XX
DT 17-AUG-1998 (first entry)

```

XX DE Cryptosporidium parvum GP900 antigen open reading frame.
XX KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection
XX KW antibody; prophylaxis; treatment; inhibition; retardation;
XX KW detection; diagnosis; human; ds.
XX OS Cryptosporidium parvum.
XX FH Key Location/Qualifiers
XX FT CDS 1..5163
XX FT /*tag= a
XX FT /product= GP900 antigen
XX FT misc_feature 524..1270
XX FT /*tag= b
XX FT /note= "region containing NINC mutations"
XX PN WO9806430-A1.
XX PD 19-FEB-1998.
XX PF 11-AUG-1997; 97WO-US14104.
XX PR 14-AUG-1996; 96US-0700651.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Gut J, Leech J, Nelson RC, Petersen C;
XX XX WPI, 1998-159290/14.
XX DR P-PSDB; W48299.
XX PT Anti-Cryptosporidium antibody - used to develop products for
XX PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
XX PT infections
XX PS Claim 32; Pages 60-62; 89pp; English.
XX CC The sequence is that encoding the GP900 antigen which may be used
XX CC in the production of anti-Cryptosporidium antibodies. These can be
XX CC used for the prophylaxis, treatment, inhibition or retardation of
XX CC a Cryptosporidium infection in humans or in animals such as calves.
XX CC They can also be used for the detection and diagnosis of related
XX CC infections.
XX SQ Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T; 0 other;

alignment_scores:
    quality: 94.50      length: 31
    ratio: 4.109      gaps: 1
    percent similarity: 74.194      percent identity: 45.161

alignment_block:
US-09-316-048-27 x V20700/rev ..

Align seg 1/1 to reverse of: V20700 from: 1 to: 5163

7 CysThrAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysGlyCysThrCy 23
||||:||||:||||:||||| ||||| ||||| ||||| |||||
865 TGTAGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 816

23 sAlaGlyAlaGlyGlyGlyThrCysThrAlaGlyThrGlyCys 37
|::||| |::||| |::||| |::|||
815 TTGTGTGT.....TGTAGTTGTGTTGTTGTTGTTGTTGTTGTTGTTG 788

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT.V20701
seq_documentation_block:
ID V20701 standard; DNA; 5318 BP.
XX AC V20701;
XX

```

```

DT 17-AUG-1998 (first entry)
XX
DE Cryptosporidium parvum GP900 antigen open reading frame and 3' region.
XX
KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW antibody; prophylaxis; treatment; inhibition; retardation;
KW detection; diagnosis; human; 3' region; ds.
XX
OS Cryptosporidium parvum.
XX
FH Key Location/Qualifiers
FT CDS 1..5166
FT FT /*tag= a
FT FT /product= GP900 antigen
FT misc_feature 524..1270
FT FT /*tag= b
FT FT /note= "region containing NINC mutations"
XX
XX WO9806430-A1.
XX PD 19-FEB-1998.
XX
XX 11-AUG-1997; 97WO-US14104.
XX PF
XX 14-AUG-1996; 96US-0700651.
XX PR
XX (REGC ) UNITV CALIFORNIA.
XX PA
XX
XX PI Gut J, Leech J, Nelson RC, Petersen C;
XX
XX WPI; 1998-159290/14.
XX DR P-PSDB; W48299.
XX
XX PT Anti-Cryptosporidium antibody - used to develop products for
XX PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
XX PT infections
XX
XX PS Claim 32; Pages 62-63; 89pp; English.
XX
XX CC The sequence is that encoding the GP900 antigen which may be used
XX CC in the production of anti-Cryptosporidium antibodies. These can be
XX CC used for the prophylaxis, treatment, inhibition or retardation of
XX CC a Cryptosporidium infection in humans or in animals such as calves.
XX CC They can also be used for the detection and diagnosis of related
XX CC infections.
XX
XX SQ Sequence 5318 BP; 1944 A; 1153 C; 888 G; 1333 T; 0 other;
XX
alignment_scores:
Quality: 94.50 Length: 31
Ratio: 4.109 Gaps: 1
Percent Similarity: 74.194 Percent Identity: 45.161
alignment_block:
US-09-316-048-27 x V20701/rev ..
Align seg 1/1 to reverse of: V20701 from: 1 to: 5318
7 CysThrAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrCys 23
|||||:|||||:||||| ||||||| |||||||:|||||
865 TGTAGTTGTGTTGTAGTTGTGTTGTGTTGTGTTGTGTTGTGTTG 816
23 salaglyAlaglyGlyGlyThrCysThrAlaGlyThrGlyCys 37
||||| |||:||||| |||:||||| :|||
815 TTGTGCT.....TGTAGTTGTGTTGTAGTTGT 788
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T80415
seq_documentation_block:
ID T80415 standard; DNA; 13987 BP.
XX
AC T80415;

```

XX 02-MAR-1998 (first entry)
DT Hybrid srmG/tylG ORF1.
XX
XX Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
KM platenolide synthase gene cluster; platenolide production; srmG gene;
KW polyketide; ty lactone synthesis; antibiotic; tylosin; hybrid gene; ss.
XX
OS Streptomyces ambofaciens.
OS Streptomyces fradiae.
XX
FH Key Location/Qualifiers
FT CDS 350..13987
FT /*tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes hybrid protein shown in W22611"
XX
PN EP791655-A2.
XX
XX 27-AUG-1997.
PD
XX 19-FEB-1997; 97EP-0301056.
PF
XX 22-FEB-1996; 96US-0012078.
PR
XX (ELIL) LILLY & CO ELI.
PA
XX Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
PI
XX WPI; 1997-418046/39.
DR P-PSDB; W22611.
DR
XX DNA encoding Streptomyces fradiae ty lactone synthase domain - for
PT production of tylosin-related polyketide compounds
PT
XX Claim 22; Pages 178-197; 220pp; English.
PS
XX This sequence represents a hybrid gene of the invention. This sequence
CC was created by replacing a EcoRI-ApaI fragment of srmG ORF1 with a
CC EcoRI-SfiI fragment from tylG ORF1. The position of the nucleotides from
CC each of the two genes is not given in the specification. The srmG gene
CC (see T80414) was isolated from Streptomyces ambofaciens, and encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The tylG gene (see T80413) is the ty lactone
CC synthase gene cluster of the invention. The tylG sequence was isolated
CC from Streptomyces fradiae, and encodes multifunctional proteins which
CC direct the synthesis of the polyketide ty lactone. Ty lactone is the basic
CC building block of the antibiotic tylosin. The hybrid sequence can be used
CC to transform S. ambofaciens lacking the srmG ORF1 sequence, or S. fradiae
CC lacking the tylG ORF1 sequence, so that they can produce polyketides. The
CC DNA sequence can be modified so as to alter the type of carboxylic acids
CC incorporated, the number of carboxylic acids incorporated and/or the
CC post-condensation reactions performed, thereby resulting in novel
CC tylosin-related polyketides.
XX
SQ Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 other;

alignment_scores:
Quality: 94.00 Length: 43
Ratio: 3.481 Gaps: 3
Percent Similarity: 62.791 Percent Identity: 46.512

alignment_block:
us-09-316-048-27 x T80415 ..

Align seg 1/1 to: T80415 from: 1 to: 13987

10 G1yAlaG1yCysG1yCysG1y...CysG1yCysThr..... 22
||||:||||||||||||| ||| |||||||:::
9787 GGGTGTGATGTGTGGTGCCTCCTGTGGGTGTCTTCGGGG 9836

23CysAlaG1yAlaG1y.....G1yG1yThrC 31
|||::||| |||||
9837 CCGGTGCTGTGTGTCGGGGAGTCCGGTGC GG GTCTCGTGGGCTGGGGGTG 9886
31 ystThrAlaG1yThrG1yCysG1yAlaG1y 40
::: |||::||| |||::|||
9887 TTGTTCAAGGGTCAAGGTGCGCACTGGGT 9915

VERSION AL022328.21 GI:5263010
KEYWORDS HTG; Cpg island; GB110; KIAA0315; KIAA0901; kinase; MAPK11; MAPK12; MHD2; Mitogen activated protein kinase; mitogen-activated; MOV10; MRS1; PRKM1; SAPK3; SPC98.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 177241)
TITLE Coville,G.
JOURNAL Direct Submission
COMMENT Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Jun 29, 1999 this sequence version replaced gi:5262835.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; SW; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP3-402G11 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-402G11 the true right end of clone RP5-89814 is at 23403 in this sequence. The true right end of clone RP11-232E17 is at 56734 in this sequence. The true right end of clone RP4-600024 is at 96013 in this sequence.
FEATURES
source location/Qualifiers
1. .177241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q13.31-13.33"
/clone="RP3-402G11"
/clone_lib="RPCI-3"
135. .447
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
657. .959
repeat_region /note="AluSg repeat: matches 1. .310 of consensus"
1373. .1659
repeat_region /note="AluSx repeat: matches 1. .298 of consensus"
1669. .1966
repeat_region /note="AluSg repeat: matches 1. .306 of consensus"
2504. .2574
repeat_region /note="MER58 repeat: matches 170. .243 of consensus"
2663. .2774
repeat_region /note="L2 repeat: matches 2643. .2747 of consensus"
2775. .3072
repeat_region /note="AluSg repeat: matches 1. .300 of consensus"
3073. .3086
repeat_region /note="L2 repeat: matches 2630. .2643 of consensus"
3959. .4455
repeat_region /note="MER1A repeat: matches 31. .527 of consensus"
5044. .5270
repeat_region /note="AluJb repeat: matches 77. .303 of consensus"
6305. .6592

misc_feature /note="AluJb repeat: matches 29. .308 of consensus"
6406. .6907
repeat_region /note="match: GSS: Em:AQ475194"
6853. .7155
repeat_region /note="AluY repeat: matches 1. .303 of consensus"
7218. .7516
repeat_region /note="AluYb8 repeat: matches 1. .318 of consensus"
9018. .9309
repeat_region /note="AluSx repeat: matches 1. .294 of consensus"
9427. .9542
repeat_region /note="AluJb repeat: matches 30. .150 of consensus"
9543. .9847
repeat_region /note="AluY repeat: matches 1. .303 of consensus"
9848. .10020
repeat_region /note="AluJb repeat: matches 150. .306 of consensus"
10024. .10316
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
10324. .10619
repeat_region /note="AluY repeat: matches 2. .297 of consensus"
10804. .11101
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
11227. .11641
repeat_region /note="MER4C repeat: matches 22. .461 of consensus"
11642. .11957
repeat_region /note="AluY repeat: matches 1. .310 of consensus"
11958. .11981
repeat_region /note="MER4C repeat: matches 1. .22 of consensus"
14063. .14370
repeat_region /note="AluY repeat: matches 2. .310 of consensus"
15488. .15718
repeat_region /note="AluSx repeat: matches 51. .295 of consensus"
16411. .16616
repeat_region /note="AluY repeat: matches 91. .296 of consensus"
16816. .17110
repeat_region /note="AluJb repeat: matches 1. .305 of consensus"
18096. .18300
repeat_region /note="MER46C repeat: matches 113. .337 of consensus"
18366. .18558
repeat_region /note="MIR repeat: matches 32. .247 of consensus"
18590. .18727
repeat_region /note="LTR45 repeat: matches 389. .525 of consensus"
21328. .21889
repeat_region /note="MER54B repeat: matches 1. .638 of consensus"
21958. .22115
repeat_region /note="L1 repeat: matches 4613. .4770 of consensus"
22116. .22245
repeat_region /note="FLAM_C repeat: matches 1. .129 of consensus"
22246. .22887
repeat_region /note="L1 repeat: matches 3975. .4613 of consensus"
22938. .23151
repeat_region /note="FRAM repeat: matches 1. .166 of consensus"
23152. .23217
repeat_region /note="L1 repeat: matches 3919. .3994 of consensus"
23218. .23373
repeat_region /note="AluSg/x repeat: matches 133. .302 of consensus"
23375. .23514
repeat_region /note="L1 repeat: matches 3780. .3926 of consensus"
23516. .23691
repeat_region /note="FRAM repeat: matches -7. .168 of consensus"
23699. .24482
repeat_region /note="L1 repeat: matches 2915. .3769 of consensus"
24501. .24747
repeat_region /note="L1MEC repeat: matches 1501. .1751 of consensus"
24978. .25194
repeat_region /note="L1MEC repeat: matches 272. .492 of consensus"
25933. .26246
repeat_region /note="AluY repeat: matches 1. .309 of consensus"
27926. .28225
repeat_region /note="AluSx repeat: matches 2. .302 of consensus"
28694. .28997
repeat_region /note="AluSg repeat: matches 1. .304 of consensus"
29001. .29119
repeat_region /note="AluJb/FLAM repeat: matches 3. .133 of consensus"

repeat_region	29859. .30170	/note="LIME repeat: matches 5189. .5489 of consensus"
repeat_region	30171. .30449	/note="AluSx repeat: matches 24. .304 of consensus"
repeat_region	30450. .30556	/note="LIME repeat: matches 5086. .5189 of consensus"
repeat_region	30567. .30813	/note="L2 repeat: matches 1079. .1303 of consensus"
repeat_region	30975. .31047	/note="LIMC/D repeat: matches 5325. .5397 of consensus"
repeat_region	31048. .31362	/note="AluY repeat: matches 1. .303 of consensus"
repeat_region	31363. .31990	/note="LIMC/D repeat: matches 4762. .5325 of consensus"
repeat_region	31997. .32351	/note="MLT1A1 repeat: matches 16. .365 of consensus"
repeat_region	32365. .32907	/note="LIMC/D repeat: matches 4184. .4755 of consensus"
repeat_region	32908. .33208	/note="AluSx repeat: matches 1. .301 of consensus"
repeat_region	33209. .33464	/note="LIMC/D repeat: matches 3936. .4184 of consensus"
repeat_region	33468. .33637	/note="FRAM repeat: matches 1. .175 of consensus"
repeat_region	33650. .33707	/note="LIMEC repeat: matches 2384. .2075 of consensus"
repeat_region	33708. .33998	/note="AluSx repeat: matches 1. .288 of consensus"
repeat_region	33999. .34144	/note="LIMEC repeat: matches 2231. .2385 of consensus"
repeat_region	34781. .34916	/note="4 copies 34 mer 86 conserved"
repeat_region	34792. .34910	/note="7 copies 17 mer 73 conserved"
repeat_region	35157. .35466	/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region	35626. .35922	/note="AluSg repeat: matches 1. .297 of consensus"
repeat_region	35932. .36042	

```
alignment_scores:
  Quality: 124.00      Length: 34
  Ratio: 5.167         Gaps: 1
  Percent Similarity: 70.588  Percent Identity: 58.824
```

alignment_block:

Align seg 1/1 to reverse of: HS402G11 from: 1 to: 177241

7 CysThrAlaGlyAlaGlyCysGlyGlyCysCys.....GlyCysGlyCy 21
||| :::::||||||| ||||| |||||
134698 TGCTGCTGTGGAACTGGGTGTGATGCTGCTGTGGACAGAGGGGTGGATG 134649

21 sthrcysalaglalyaglygcytythr cysthr alaglythr gcy sc 38
| : : : : | | | | : : : | : : : |
134648 CTCCTGCTGTGGACGGGGTGTGGCGCATGCTGCTGTGGACCTGCATCTG 134599

38 1y 38
134598 GA 134597

seq_name: gb_in2:BMOCHRCA

```

seq_documentation_block:
LOCUS      BMOCHRHCA      2182 bp      DNA
DEFINITION B.mori (silkmoth) high-cysteine chorion gene pair Hc-A.12/Hc-B.12,
            complete cds.

```

ACCESSION	K02835
VERSION	K02835.1
KEYWORDS	GI:155985 silmoth protein.
SOURCE	Silkmoth DNA, clones Hc-A.12 and Hc-B.12.
ORGANISM	Bombyx mori

REFERENCE AUTHORS TITLE	JOURNAL MEDLINE COMMENT
Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Bombycidae; Bombyx. 1 (bases 1 to 2182) Iatrou,K., Tsitliou,S.G. and Kafatos,F.C. DNA sequence transfer between two high-cysteine chorion gene families in the silkmoth Bombyx mori Proc. Natl. Acad. Sci. U.S.A. 81, 4452-4456 (1984)	84272653 Putative TATA boxes are at positions 1154-1160 for Hc-A.12 and 1370-1376 for Hc-B.12. Putative polyadenylation signals are located at positions 201-206 and 1977-1982 for Hc-A.12 and Hc-B.12 respectively.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 2182)	Iatrou, K., Tsitilou, S.G. and Kafatos, F.C.	DNA sequence transfer between two high-cysteine chorion gene families in the silkworm <i>Bombyx mori</i>	
		Proc. Natl. Acad. Sci. U.S.A. 81, 4452-4456 (1984)		
		84272653		
		Putative TATA boxes are at positions 1154-1160 for Hc-A.12 a		

Putative TATA boxes are at positions 1154-1160 for Hc-A.12 and 1370-1376 for Hc-B.12. Putative polyadenylation signals are located at positions 201-206 and 1977-1982 for Hc-A.12 and Hc-B.12, respectively.

FEATURES	Location/Qualifiers
----------	---------------------

Source

CDS

```
Location/Qualifiers
1. 2182
/organism="Bombyx mori"
/db_xref="taxon:7091"
Complement(join(241. 564,1047. 1097))
```

```

/codon_start=1

```

```

/protein_id="AAA27834.1"

```

```
/translation="MTFALLLCVQGLIQNVYGCCGGCGGGCGGCGCGYGGEDGN
```

```
complement(<241, .564)
```

```

/...:  ...:  ...:
/number=2

```

```
complement(244.0002)
/product="high cysteine chorion A"
```

```
complement(join(553: :564,1047: :1057))
/note="signal peptide for high-cysteine-chorion A"
```

```
complement(500:1040)
/note="Ch-A"
```

```
complement(1047..1127)
note="Ch-A"
```

```

/number=1
join(1427. .1477,1559. .1903)

```

```
/note="precursor"  
/codon start=1
```

```
/product="high cysteine chorion B"
/protein_id="AAA27835.1"
```

```
/db_xret="GI:155987"
```

NRVCSNSAAPGTGLSICSENNRYKGDVCVCCEVPFLGTADVCGNMCMSSGGCCIDYGC

```
join(1427, .1477, 1559, .1570)
```

1478. .1558

1559. >1903

```

/number=2

```

```

15741: 1.1500
/product="high cysteine chorion B"

```

stream of HinfI site.

alignment_scores:

Quality:	122.50	Length:	45
Ratio:	3.952	Gaps:	2
Percent Similarity:	68.889	Percent Identity:	51.111

alignment_block:

US-09-316-048-27 x BMOCHRCA/rev ..

Align seg 1/1 to reverse of: BMOCHRHCA from: 1 to: 2182

2 AlaCysAlaAlaThrcysthAla.....GlyAlaGl 12
 |||||::: ||::: ||:::||
 393 GCCTGTGCTAGTGGTTGTGTCTCTATCTGCGACGCTGCTGCGGATGCGG 344

*	13907	14006:	gap of unknown length
*	14007	15090:	contig of 1084 bp in length
*	15091	15190:	gap of unknown length
*	15191	16431:	contig of 1241 bp in length
*	16432	16531:	gap of unknown length
*	16532	17756:	contig of 1225 bp in length
*	17757	17856:	gap of unknown length
*	17857	19038:	contig of 1182 bp in length
*	19039	19138:	gap of unknown length
*	19139	20372:	contig of 1234 bp in length
*	20373	20472:	gap of unknown length
*	20473	21712:	contig of 1240 bp in length
*	21713	21812:	gap of unknown length
*	21813	22979:	contig of 1167 bp in length
*	22980	23079:	gap of unknown length
*	23080	24142:	contig of 1063 bp in length
*	24143	24242:	gap of unknown length
*	24243	25417:	contig of 1175 bp in length
*	25418	25517:	gap of unknown length
*	25518	26565:	contig of 1048 bp in length
*	26566	26665:	gap of unknown length
*	26666	27713:	contig of 1048 bp in length
*	27714	27813:	gap of unknown length
*	27814	28946:	contig of 1133 bp in length
*	28947	29046:	gap of unknown length
*	29047	30187:	contig of 1141 bp in length
*	30188	30287:	gap of unknown length
*	30288	31532:	contig of 1245 bp in length
*	31533	31632:	gap of unknown length
*	31633	32859:	contig of 1227 bp in length
*	32860	32959:	gap of unknown length
*	32960	34118:	contig of 1159 bp in length
*	34119	34218:	gap of unknown length
*	34219	35277:	contig of 1059 bp in length
*	35278	35377:	gap of unknown length
*	35378	36613:	contig of 1236 bp in length
*	36614	36713:	gap of unknown length
*	36714	37965:	contig of 1252 bp in length
*	37966	38065:	gap of unknown length
*	38066	39167:	contig of 1102 bp in length
*	39168	39267:	gap of unknown length
*	39268	40333:	contig of 1066 bp in length
*	40334	40433:	gap of unknown length
*	40434	41793:	contig of 1360 bp in length
*	41794	41893:	gap of unknown length
*	41894	44026:	contig of 2133 bp in length
*	44027	44126:	gap of unknown length
*	44127	45784:	contig of 1658 bp in length
*	45785	45884:	gap of unknown length
*	45885	47475:	contig of 1591 bp in length
*	47476	47575:	gap of unknown length
*	47576	49524:	contig of 1949 bp in length
*	49525	49624:	gap of unknown length
*	49625	52194:	contig of 2570 bp in length
*	52195	52294:	gap of unknown length
*	52295	56081:	contig of 3787 bp in length
*	56082	56181:	gap of unknown length
*	56182	60752:	contig of 4571 bp in length
*	60753	60852:	gap of unknown length
*	60853	64243:	contig of 3391 bp in length
*	64244	64343:	gap of unknown length
*	64344	72855:	contig of 8512 bp in length
*	72856	72955:	gap of unknown length
*	72956	79691:	contig of 6736 bp in length
*	79692	79791:	gap of unknown length
*	79792	88438:	contig of 8647 bp in length
*	88439	88538:	gap of unknown length
*	88539	95200:	contig of 6662 bp in length
*	95201	95300:	gap of unknown length
*	95301	104456:	contig of 9156 bp in length
*	104457	104556:	gap of unknown length
*	104557	113122:	contig of 8566 bp in length
*	113123	113222:	gap of unknown length

FEATURES	source
*	113223 131437: contig of 18215 bp in length
*	131438 131537: gap of unknown length
*	131538 147555: contig of 16018 bp in length
*	147556 147555: gap of unknown length
*	147656 183113: contig of 35458 bp in length
*	183114 183213: gap of unknown length
*	183214 210273: contig of 27060 bp in length
*	210274 210373: gap of unknown length
*	210374 273658: contig of 63285 bp in length.
	Location/Qualifiers
	1. .273658
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/chromosome="16"
	/clone="RP23-10A2"
	/clone_lib="RPCI mouse BAC library 23"
BASE COUNT	74849 a 54887 c 68561 g 60594 t 14767 others
ORIGIN	

alignment_scores:		
Quality:	119.50	Length: 48
Ratio:	3.621	Gaps: 2
Percent Similarity:	68.750	Percent Identity: 43.750

alignment_block:
US-09-316-048-27 x AC079423/rev ...

Align seg 1/1 to reverse of: AC079423 from: 1 to: 273658

1 ThrAlacysAlaIatHrCysThrAlaGlyAlaGlyCysGlyGlyCysCy 17
 :::::||||::: ||| ::|||::| ||||| |||||
 125408 TCtACCTGCGAGGAGGCTGC...TGCGGCTCTGGCTGCGGGGATGCGG 125362

17 scgIcysgIcysthrCysAlaGlyAlaGly..... 27
 |||||
 ::|||::|||::|||
 125361 AGGCTGCCGGGTCTAGCTGCTGCCGATCTGGCTGCCGAGGAGGCTGCCGG 125312

```

28 .....GlyGlyThrCysThrAlaGlyThrGlyCysGlyAlaGly 40
      ||||| ||| ::|||:::||||| ::|||
125311 GCTGTGGAAGCGCGCTGCTGTGGCGGCTCCGGATGCTGCGGTGT 125268

```

seq_name: gb_htg24:LMFLCHR12_07

seq_documentation_block:

Sequence split into 23 fragments LOCUS LMF1CHR12 Accession AL390114

LMFLCHR12_00

LMF1CHR12_02	200001	310000
LMF1CHR12_03	300001	410000
LMF1CHR12_04	400001	510000
LMF1CHR12_05	500001	610000
LMF1CHR12_06	600001	710000
LMF1CHR12_07	700001	810000
LMF1CHR12_08	800001	910000
LMF1CHR12_09	900001	1010000
LMF1CHR12_10	1000001	1110000
LMF1CHR12_11	1100001	1210000
LMF1CHR12_12	1200001	1310000
LMF1CHR12_13	1300001	1410000
LMF1CHR12_14	1400001	1510000
LMF1CHR12_15	1500001	1610000
LMF1CHR12_16	1600001	1710000
LMF1CHR12_17	1700001	1810000
LMF1CHR12_18	1800001	1910000
LMF1CHR12_19	1900001	2010000
LMF1CHR12_20	2000001	2110000
LMF1CHR12_21	2100001	2210000
LMF1CHR12_22	2200001	2287483

Continuation (8 of 23) of LMFCHR12 from base 700001 (AL390114 *Leishmania major* chrom

```

alignment_scores:      quality: 119.00      length: 39
                       ratio: 4.103      gaps: 2
Percent Similarity: 74.359      Percent Identity: 56.410

alignment_block:
US-09-316-048-27 x LMFLCHR12_07/rev ..

Align seg 1/1 to reverse of: LMFLCHR12_07 from: 1 to: 110000

      2 AlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysCysGly 18
      |||||:::  |||  ::|||:::|||||  |||||
34465 GCCTCGTGTGTGAGCGTGTGCGTGTGNTGTGGGTGTGGGTGTGGGTGTGG 34416

      18 yCysGlyCysThrCysAlaGlyAlaGlyGlyThrCysThrAlaGlyT 35
      |||||  |||  |||:::  |||  |||  :::  |||
34415 GTGTGGGTGTGGGTGT...GGGTGTGGGTGTGG...TGTGGGTGTGGGT 34372

      35 hrglyCysGlyAlaGly 40
      |||||:::  |||
34371 GTGGGTGTGGGTGCCGA 34355

seq_name: gb_htg17:AC073826

seq_documentation_block:
LOCUS      AC073826      146190 bp      DNA      HTG      29-JUN-2000
DEFINITION Mus musculus clone RP23-93K16, WORKING DRAFT SEQUENCE, 35 unordered
            pieces.
ACCESSION  AC073826
VERSION    AC073826.1 GI:8810443
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.

REFERENCE  1 (bases 1 to 146190)
AUTHORS   DOE Joint Genome Institute.
TITLE     Sequencing of Mouse
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 146190)
AUTHORS   DOE Joint Genome Institute.
TITLE     Direct Submission
JOURNAL    Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1772840
Center clone name: RPCI-23_93K16
-----
Summary Statistics
Consensus quality: 118979 bases at least. Q40
Consensus quality: 131063 bases at least Q30
Consensus quality: 133821 bases at least Q20
Estimated insert size: 187180; agarose-fp estimation
Estimated insert size: 142790; sum-of-contigs estimation
Quality coverage: 8.66 in Q20 bases; agarose-fp estimation
Quality coverage: 11.35 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1067: contig of 1067 bp in length
* 1068 1167: gap of unknown length
* 1168 2519: contig of 1352 bp in length

```

FEATURES	source	1. .146190	/organism="Mus musculus"
		/db_xref="taxon:10090"	
		/clone="Rp23-93K16"	
		/clone_1fb="RPC1 mouse BAC library 23"	
BASE COUNT	35412 a	36979 c	36009 g 34377 t 3413 others

ORIGIN

alignment_scores: Quality: 118.00 Length: 39
 Ratio: 4.214 Gaps: 0
Percent Similarity: 71.795 Percent Identity: 48.718

alignment_block:
US-09-316-048-27 x AC073826/rev ..

Align seg 1/1 to reverse of: AC073826 from: 1 to: 146190

2 AlAcysAlaIaIaThrCysThraIaGlyAlaGlyCysGlyCysG1 18
:::||||::: ||| ::|||:::||||||| |||||
139701 AGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 139652

18 yCysGlyCysThrCysAlaGlyAlaGlyGlyThrCysThraIaGlyT 35
||||||| |||::: ||| ||| ::|||
139651 CTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCT 139602

35 hrGlyCysGlyAlaGly 40
|||||||::: |||

139601 GTGGCTGTGGCTGTGGC 139585

seq_name: gb_htg17:AC073748

seq_documentation_block:
LOCUS AC073748 175371 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-305J2, WORKING DRAFT SEQUENCE, 33 unordered
pieces.
ACCESSION AC073748
VERSION AC073748.1 GI:8810365
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 175371)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175371)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1854210
Center clone name: RPCI-23_305J2

Summary Statistics
Consensus quality: 152503 bases at least Q40
Consensus quality: 164435 bases at least Q30
Consensus quality: 166779 bases at least Q20
Estimated insert size: 166550; agarose-fp estimation
Estimated insert size: 172171; sum-of-contigs estimation
Quality coverage: 11.93 in Q20 bases; agarose-fp estimation
Quality coverage: 11.54 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1182: contig of 1182 bp in length

* 1183 1282: gap of unknown length
* 1283 2853: contig of 1571 bp in length
* 2854 2953: gap of unknown length
* 2954 4183: contig of 1230 bp in length
* 4184 4283: gap of unknown length
* 4284 5786: contig of 1503 bp in length
* 5787 5886: gap of unknown length
* 5887 7066: contig of 1180 bp in length
* 7067 7166: gap of unknown length
* 7167 8238: contig of 1072 bp in length
* 8239 8338: gap of unknown length
* 8339 9718: contig of 1380 bp in length
* 9719 9818: gap of unknown length
* 9819 11510: contig of 1692 bp in length
* 11511 11610: gap of unknown length
* 11611 12959: contig of 1349 bp in length
* 12960 13059: gap of unknown length
* 13060 14954: contig of 1895 bp in length
* 14955 15054: gap of unknown length
* 15055 16862: contig of 1808 bp in length
* 16863 16962: gap of unknown length
* 16963 19187: contig of 2225 bp in length
* 19188 19287: gap of unknown length
* 19288 21631: contig of 2344 bp in length
* 21632 21731: gap of unknown length
* 21732 23582: contig of 1851 bp in length
* 23583 23682: gap of unknown length
* 23683 25787: contig of 2105 bp in length
* 25788 25887: gap of unknown length
* 25888 27450: contig of 1563 bp in length
* 27451 27550: gap of unknown length
* 27551 31646: contig of 4096 bp in length
* 31647 31746: gap of unknown length
* 31747 33311: contig of 1565 bp in length
* 33312 33411: gap of unknown length
* 33412 36480: contig of 3069 bp in length
* 36481 36580: gap of unknown length
* 36581 40140: contig of 3560 bp in length
* 40141 40240: gap of unknown length
* 40241 46716: contig of 6476 bp in length
* 46717 46816: gap of unknown length
* 46817 52150: contig of 5334 bp in length
* 52151 52250: gap of unknown length
* 52251 56774: contig of 4524 bp in length
* 56775 56874: gap of unknown length
* 56875 62585: contig of 5711 bp in length
* 62586 62685: gap of unknown length
* 62686 72379: contig of 9694 bp in length
* 72380 72479: gap of unknown length
* 72480 78605: contig of 6126 bp in length
* 78606 78705: gap of unknown length
* 78706 87900: contig of 9195 bp in length
* 87901 88000: gap of unknown length
* 88001 98832: contig of 10832 bp in length
* 98833 98932: gap of unknown length
* 98933 109561: contig of 10629 bp in length
* 109562 109661: gap of unknown length
* 109662 127036: contig of 17375 bp in length
* 127037 127136: gap of unknown length
* 127137 139467: contig of 12331 bp in length
* 139468 139567: gap of unknown length
* 139568 158423: contig of 18856 bp in length
* 158424 158523: gap of unknown length
* 158524 175371: contig of 16848 bp in length.
Location/Qualifiers
1. 175371
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-305J2"
/clone_lib="RPCI mouse BAC library 23"
44046 a 42634 c 42369 g 43110 t 3212 others

BASE COUNT
ORIGIN

```

alignment_scores:
  Quality: 116.00      Length: 39
  Ratio: 4.000        Gaps: 2
  Percent Similarity: 74.359      Percent Identity: 53.846

alignment_block:
  US-09-316-048-27 x AC073748 ..

Align seg 1/1 to: AC073748 from: 1 to: 175371

      2 AlacysAlaAlaThrcysThraAlaGlyAlaGlyCysGlyCysCysG1 18
      ::::::::::: ||| ::::::::::: |||
32902 AGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 32951
      18 yCysGlyCysThrcysAlaGlyAlaGlyGlyThrcysThraAlaGlyT 35
      ||| ||| ||| ||| ||| ||| ||| |||
32952 CTGTGGCTGTGGCTGT...GGCTGTGGCTGTGGGC...TGTGCTGTGGCT 32995

      35 hrglyCysGlyAlaGly 40
      ||| ||| ||| ||| ||| ||| ||| |||
32996 GTGGCTGTGGCTGTGGC 33012

seq_name: gb_htg17:AC073748

seq_documentation_block:
LOCUS AC073748 175371 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-305J2, WORKING DRAFT SEQUENCE, 33 unordered
pieces.
AC073748
AC073748.1 GI:8810365
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 175371)
DOE Joint Genome Institute.
TITLE Sequencing of Mouse
Unpublished
2 (bases 1 to 175371)
DOE Joint Genome Institute.
JOURNAL Direct Submission
REFERENCE Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE -----Genome Center
JOURNAL Center: Joint Genome Institute
AUTHORS Center Code: JGI
JOURNAL Web site: http://www.jgi.doe.gov
COMMENT
-----
Project Information
Center Project Name: 1854210
Center clone name: RPCI-23_305J2
-----
Summary Statistics
Consensus quality: 152503 bases at least Q40
Consensus quality: 164435 bases at least Q30
Consensus quality: 166779 bases at least Q20
Estimated insert size: 166550; agarose-fp estimation
Estimated insert size: 172171; sum-of-contigs estimation
Quality coverage: 11.93 in Q20 bases; agarose-fp estimation
Quality coverage: 11.54 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
      1 1182: contig of 1182 bp in length
      * 1183 1282: gap of unknown length
      * 1283 2853: contig of 1571 bp in length

```

FEATURES	source	1. .175371	/organism="Mus musculus"	
			/db_xref="taxon:10090"	
			/clone="RP23-305J2"	
			/clone_lib="RPCI mouse BAC library 23"	
BASE COUNT	ORIGIN	44046	a 42634	c 42369
			g 43110	t 3212
				others

alignment_scores:

JOURNAL Infect. Immun. 68 (8), 4815-4817 (2000)
MEDLINE 20359389
REFERENCE 2 (bases 1 to 323)
AUTHORS Boursaux-Eude, C.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-1999) Boursaux-Eude C., Laboratoire des
Bordetella, Institut Pasteur, 25 rue du Dr Roux, Paris, 75015,
FRANCE

FEATURES
source

1. .323
/organism="Bordetella bronchiseptica"
/isolate="SEI"
/db_xref="taxon:518"
/country="France"
1. .323
/gene="prn"
<1. .>323
/gene="prn"
/note="in the (POP)n repeat region (PRN10-2)"
/codon_start=2
/transl_table=11
/product="pertactin (P.68)"
/protein_id="CAB76449.1"
/db_xref="GI:7160666"
/translation="GTYRYRLAANGNGOWSLYGAKVPAPKPAPOPPOPPOPPO
PPQPQPQPQPPEAPAPQPAPAGRELSAANAAVNTGGVGLASTLWYAESNALSKRLGE
LRLNP"
CDS

BASE COUNT 49 a 124 c 119 g 31 t
ORIGIN

alignment_scores:
Quality: 112.50 Length: 40
Ratio: 3.879 Gaps: 4
Percent Similarity: 72.500 Percent Identity: 62.500

alignment_block:
US-09-316-048-27 x BBR250095/rev ..
Align seg 1/1 to reverse of: BBR250095 from: 1 to: 323

4 AlaAlaThrCysThrAlaGlyAla...GlyCysGlyGlyCysGlyCys 19
||||: ||| ||||||||| ||||||||| |||||||||
195 GCCGGCGGTTGCGGCGCGCGCTTCGGGCTGCGGC...TGGCGCTG 152
19 sGlyCysThrCysAlaGlyAlaGlyGly...GlyThrCysThrAla...G 34
||||||| |||:||||:||||||| ||| ||| ::|
151 CGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 102
34 lYThrGlyCysGlyAlaGly 40
|| |||||||||||||
101 GACCCGCGCTGCGGCGCGGC 82

seq_name: gb_in2:BMHCBPGP

seq_documentation_block:

LOCUS BMHCBPGP 418 bp mRNA INV 19-JUL-1995
DEFINITION Silkmoth mRNA for transcribed pseudogene Hcb.12/13 (partial).
ACCESSION X15513
VERSION X15513.1 GI:5880
KEYWORDS chorion protein; pseudogene.
SOURCE domestic silkworm.
ORGANISM Bombyx mori

REFERENCE 1 (bases 1 to 418)
AUTHORS Fotaki,M.E. and Iatrou,K.
TITLE Identification of a transcriptionally active pseudogene in the
chorion locus of the silkmoth Bombyx mori. Regional sequence
conservation and biological function
JOURNAL J. Mol. Biol. 203 (4), 849-860 (1988)
MEDLINE 89094862

COMMENT See also <X13860> for another B.mori HCB pseudogene.

FEATURES
source

Location/Qualifiers
1. .418
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone="m2282"
<1. .254
/note="pot. chorion protein (84 AA) (1 is 2nd base in
codon)"
/pseudo
/codon_start=3
/db_xref="PID:e189283"
99. .101
/note="internal stop codon"
287. .292
/note="pot. polyA signal"
401. .406
/note="pot. alt. polyA signal"
418
/note="polyA site"

BASE COUNT 108 a 55 c 143 g 112 t
ORIGIN

alignment_scores:
Quality: 112.50 Length: 29
Ratio: 5.114 Gaps: 1
Percent Similarity: 75.862 Percent Identity: 65.517

alignment_block:
US-09-316-048-27 x BMHCBPGP ..
Align seg 1/1 to: BMHCBPGP from: 1 to: 418

10 GlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrCysAlaGlyAl 26
||||:||||||| ||||||||| ||| |||:
105 GGATGCGGTTGTGAGAGTGTGAGATGTGATGCTGTGTAGAGCTTG 154
26 aGlyGlyGlyThrCysThrAlaGlyThrGlyCysGly 38
:||||||| ||| ::||| |||||||||
155 CGGTGATGCGGATGCGGCTGTGA...GGTGTGGA 188

seq_name: gb_in2:BMHCBPGN

seq_documentation_block:

LOCUS BMHCBPGN 1591 bp DNA INV 22-NOV-1993
DEFINITION Silkmoth chorion high cysteine protein (Hcb) pseudogene.
ACCESSION X13860
VERSION X13860.1 GI:5879
KEYWORDS chorion protein; Hcb pseudogene; pseudogene.
SOURCE domestic silkworm.
ORGANISM Bombyx mori

REFERENCE 1 (bases 1 to 1591)
AUTHORS Fotaki,M.E. and Iatrou,K.
TITLE Identification of a transcriptionally active pseudogene in the
chorion locus of the silkmoth Bombyx mori. Regional sequence
conservation and biological function
JOURNAL J. Mol. Biol. 203 (4), 849-860 (1988)
MEDLINE 89094862

COMMENT See also <X15513> for Hc mRNA 3' flank.

FEATURES
source

Location/Qualifiers
1. .1591
/organism="Bombyx mori"
/db_xref="taxon:7091"
/clone_lib="PBR322"
/clone="sc6/150"
17. .26
/note="consensus sequence"
31. .40
/note="consensus sequence"
42. .51
/note="consensus sequence"

/product="protein tyrosine phosphatase 2"
/protein_id="AAC59008.1"
/db_xref="GI:1911255"
/translation="MYEASQIDEHVEVGGYGDNEAMLREIENHAIASVISLIDSDVA
PIRHALGLPVGDHIHVCEAAPTCALPNAMPALYDYVVRRIEGEKRVLVHCYAGASR
SAALAVYYLMRSROMPYQDALNVAQSKRVAINDHFVRLATRCYSYRVNDELKPQIV
"
complement(6129..6791)
/note="ORF10; ptp-1, AcMNPV ORF1 homolog"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAC59009.1"
/db_xref="GI:1911256"
/translation="MLTSRMAVIYTRTSERNCGAMCTNGVRRRQVHLPSVARAM
FPDRWHEYTACGAVIEGTRLLCEKRVPLNABELFEYVTSDEDRWTAASVLARHSALGAVI
DLTNTARYYDGAQMVKMGLYKKIRVPGRAVPDDDIVAEFIETVDEFFRRCPMTLVAV
HWTGHLNRSGLVLCRYMVERLGSPTDAIARFETARGHKIERITNYLQDLLARKHVRGO
PN"
6819..7814
/note="ORF11; AcMNPV ORF11 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59010.1"
/db_xref="GI:1911257"
/translation="MSLSKLLYAYVYGSYNLPHDRYGESYHLIRIVHEHTTNYVSN
ASCVRRDIATARCLNSGLCFDVAQQLDVSEVAARLSAMFERCGDATGLCADMQBALA
DIDRHAPLARVRGRANIFALDIADIPSDVTNNLGIIGREMHFPRCSGLARVADVE
DPDIRADGWWYHKFCVLTMYHLVAGAVPAGSATRLRDVAKHIGPNDGNCAPAI
VYGRFCAIGREHFAHKTACMHILFQFMNDLTPADERHPCFVGIKDEGRCKDPTD
LRTHADALYIHGTTDRQKNALFDLCCVNASDIDADCYDCVNVNKFYATQNKKKYKM"
complement(8180..9142)
/note="ORF12; AcMNPV ORF12 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59011.1"
/db_xref="GI:1911258"
/translation="MFSWMFGWNADEQVNAEFDEQAYRRYAVDQRAHSDLVRDVE
RCHPFVFKFRYVLDETAGRCSSVDFCKGLKISHDLQRCNFDROHVRLNELVIGAP
PAQPDLSGLFATKHGLVQLQOFSFANKNEVLAVGANKDNDNLDKIEAVLNHV
KTLNTNSDKFISAHKSKLEVGAPEQOEQRLQTLDTKLNALQCAAPTRTAPGVFPR
DVTKHPLAVFMGRVEDRGVTQJAFARGQEEHFRKKLFEEGMDVDVGRAPNPPLA
VHCIEEPANGGHKIRRLPKKIVIEYDCAVNAVAKDIVKKAIIINKT"
complement(9094..9825)
/note="ORF13; lcf-1, AcMNPV ORF14 homolog"
/codon_start=1
/product="late expression factor 1"
/protein_id="AAC59012.1"
/db_xref="GI:1911259"
/translation="MAPCKYTPERYQMMDAIAYNDSRLAFMTDRPRWHAHNFPDS
AAQLFAYIVKNSISDVHKPLEEGGREWVIDADFKDCAKAEMLKVNVGATAFMLEF
EGKEDAVQRIWMSGNRGFLWLKFCGKEMDAPKSLREHWNVFKQPAKLVSQDIRPG
SFADCVRAVHMYIGDAREDLVLRWPDVDVDFVCNANKQIRAPFSYNYKGGYSRCL
TQOLQORIKACSAAGLAGTPTPSK"
9893..11362
/note="ORF14; egt, AcMNPV ORF15 homolog"
/codon_start=1
/product="ecdysteroid UDP-glucosyltransferase"
/protein_id="AAC59013.1"
/db_xref="GI:1911260"
/translation="MVFLIALTLATGARAASILAVLPTPAYSHHVYRAYVHALVK
NCHNVTVIKPOLLDYAVQDECGRVEQIDADMSAQYKKLVASSGVFRKRGVADETV
TADNYMGLIEMFKDQFDNANVRFLSTNRTFDVAVVEAFADYALVFGHLRPAPVIOI
APGYGLAENEFERRAVARHPLHYPTFGAALTRRGALSEMRLLNEFELLARSDLL
KQOFGKSTPTIRQLRDNVOLLNLNHPYDNNRPVPPSVQYLGGLHLAQALPORLDA
PLERLINESVDGAVVYSESGIDTNSIHAFLQMLDITFANLNNTYVLWKVDDAVAAS
VALPRNVLAQWESQTAVALRHKNVAVFTQAQSSDEALQARVPMVCLPMGDOFHH
ARKLQGFVARALDTPAASAPQLQLAIREVIADGEAYRARIDKLRAVEVHDAAPDEKA
VKFERVIRKFNNDVNMPPARSLKTTAANMAYSDYFVRPL"
11547..12140
/note="ORF15; AcMNPV ORF16 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59014.1"

/db_xref="GI:1911261"
/translation="METAQPPISYAPPKRGAVCAVVRVVTYTTVSDSGGNEDRLTQ
IVAQLQRTKRLNFSKLSQLQRRVRNMOKLIRKNSVIANLAARLTQKTKHFAVTIR
KNVIHTTSGSEQEVRORLLELCANGSEQVCARADCARDRRRVAEALATALGAGVVA
SAANKRFEIEDEEKLVSAKLIVQVYLHDGDHSDTCAD"
12109..12732
/note="ORF16; AcMNPV ORF17 homolog"
/codon_start=1
alignment_scores:
Quality: 112.50 Length: 48
Ratio: 3.879 Gaps: 3
Percent Similarity: 60.417 Percent Identity: 45.833
alignment_block:
US-09-316-048-27 x OPU75930/rev ..
Align seg 1/1 to reverse of: OPU75930 from: 1 to: 131993
4 AlaAlaThrCysThrAlaGly.....AlaGlyCysGlyGlyCysCys.. 17
:::||||:::|||| :::|||||:|||||:|||||
58315 AGCGCCACTGTGTGCGCGCGGCTGCGGCGCTGCGGCGGCTGCTGCGG 58266
18GlyCysGlyCysThrCysAlaGlyAlaGlyGlyGlyThrc 31
|||||:||||| ||:::||||::: ||||| 1
58265 CGGCTGCGGCGGCTGCGGCTGCTGCTGCGGCGGCTGCTGCGGCGGCTGCT 58216
31 ys.....ThrAlaGlyThrGlyCysGly 38
|| :::||| |||||
58215 GCGGCGGCTGCGGCTGTTGGGTTGCTGCGGAGGTGCTGCGGCG 58172

```

;
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 95:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1904 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: THPIAZS08
; CLONE: 2754573
;
US-08-933-750C-95

alignment_scores:
      Quality: 212.00      Length: 499
      Ratio: 1.044      Gaps: 16
Percent Similarity: 40.681      Percent Identity: 22.244

alignment_block:
US-09-316-048-11 x US-08-933-750C-95 ..

Align seg 1/1 to: US-08-933-750C-95 from: 1 to: 1904

46 GlyGlyGlnGlnPheAsnGlyLysGlyLysMetArgProProValLysHi 62
   |||||
17 GGTGGCGGGCGCTGGANNTCCCTCCGGGGAGCACAGCGACCGGAGCATGCA 66
62 sAspAlaProAlaArgAlaLeuArgThrLeuGlnSerGluProLeuPro. 78
   |:::
67 CGTCGTCGACCAAGCTTCACCTGAGGCTTGAACAGGAAGTAATCTCCCTC 116
78 .....
117 CATCTCCAACCTTGCTTGACAAGCTTCCTTCTCCAGCGCGTGAGCCG 166
79 .....GluArgLeuGluValAsnGlyArg 86
167 TCCCTTCCTCCCATGTGCCAGAGCAGGACCGCGCGGAGCCCGCCGCT 216
86 gGluAlaGlnGlnGluAsnLeu..... 93
217 GGAGCGCGAGGACAGTGGACCTTCAGTTTGGGAAGATGATAACAGCTA 266
94 ....AsnLysSerGluIleSerGlnValPheGluIleAlaLeuLys... 107
   ::|||::: |||::: :: ||| ::::|||
267 AGCCAGGGAACACACCGATTTCAGTATTACACGAATACGGCATGAAGACC 316
108 ArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProHi 124
   ::|||:::||||| ::||| |||:::
317 AAGAACAATCCACGTT..TATGAATGTGAAGAATCTGATGTGCAATACA 363
124 smetLysAsnPheValThrArgValSerValGlyGluPheValGlyGlu 141
   |::: ::|||::: |||||:::|||||:::
364 CGTGCCCACTTTCACCTTCAGAGTAACCGTTGGTGCATTAACCTGCACAG 413
141 lyGlyLysSerLysLysIleSerLysLysAsnAlaAlaArgAlaVal 157
   ||||| ||||| |||||::: ||| ::|||::: |||||
414 GTGAAGGTACAAGTAAGAAGCTGGCGAAACATAGAGCTGCAGAGGCTGCC 463
158 LeuGlnLeuArgArgLeuProProLeuProAlaValGluArgVally 174
   ::
464 ATA..... 466
174 sProArgIleLysLysSerGlnProThrCysLysThrAlaProAsp. 190
   ::||| ||| ::| ||| ::::|||
467 ....AACATTTTGAAGCAAGCATGCAAGTATTGCTTTGACAGTTCTGACC 512
191 .....TyrglyGlnGlyMetAsnProIle 198
```

```

513 CCTTAATGCGCTGACCCCTTCCAAAGCAACCAAGAACACAGCTTAATCTATT 562
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTy 215
   :: ||| ::::: :: |||||
563 GGTTCATTACAGGAATTGGCTATTTCATCATGGCTGGAGACTTCCGATA 612
215 rmetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
   | ||| ||| ||| ::::|||:::
613 TACCCTTTCCAGGAGGAGGAGGACCTGCTCATAGAAGAGAAATACTACA 662
232 InValLysValGlyHisIsthrAlaGluGlyValGlyThrAsnLysLys 248
   ::::: ||| ||| |||:::
663 TTTGCAGGCTAGAGTCATTATGGAACCTGGAAGGGGCGCATCAAAAAG 712
249 ValAlaLysArgAsnAlaAlaGluAsnmetLeuGluIleLeuGlyPheLy 265
   ||||| ||||| |||||::: |||
713 CAAGCCAAAGGAATGCTGCTGAGAAATTCTT..... 745
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
745 ..... 745
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
   :: ||| :::
746 .....GCCAAATTAGTAATATT 763
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
   ||||| ||||| |||
764 TCTCA.....GAGAACACATTTCT..... 784
315 tProTyLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
784 ..... 784
332 roGluValAlaGlnAlaValAlaGlyValSerGlnGly.....HisHis 345
   ::::: ||||| ||| |||
785 ....TTAACAAATGTAGTAGACATCTTTAGATGTACTGGCATTCC 829
346 ThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValTh 362
   :::::
830 TTGAGGAAT..... 838
362 rAlaMetIleAlaArgGluLeuLeuTyrglyGlyThrSerProThrAlaG 379
839 .....TCTCTGGTGAA 851
379 lu..ThrIleLeuLysSerAsnIleSerSerGlyHisValProHisGly 394
   :: ::::: ||||| ::::: |||
852 AGATCAACTTACTGAAAGAACGCTCTAGT.....ATT 886
395 ProArgThrArgProSerGluGlnLeuTyrglyLeuSerArgAlaGlnI 411
   |||::: ||| ::| ::::: |||||
887 CCAATATACAGATTACATCCAGCTGCTAGTGAATTTGCCAAGCAACAAG 936
411 yPheGlnValGluTyrglyLysAspPheProLys.....AsnAsnLysA 425
   |||||::: ||| |||::: ::|
937 TTTTATATATACATATTGGATATATGATGACAGCCCAATGGAACAAT 986
987 ATCAATGT.....CTTGCTGAACCTGCCACAGCCCATTCACAGTCTGT 1030
425 snGluCysValSerLeuIleAsnCysSerSerGlnProProLeuValSer 441
   ::||| ||| ::| |||::: |||
442 HisGlyIleGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAs 458
   ||||| ||| ::| ::::: ||||| |||
1031 CATGGCTCGGTATCTCCTGTGGCAATGCACAAGTATGACGCTCACAA 1080
458 nIleLeuLysLeu.....Serg 465
   | |||::: |||
1081 TGCTTTGCAGTATTAAAGATATAGCAGAAAGAAAGTAAATCTGAGACA 1130
465 luleuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyPro 480
   ::||| ::| |||::: |||||
```



```

1992 AGAGCCCCGTACACACTGCTGTAGTGTATGCAC..AAATTGGGGAAC 2038
212 GluProgluTyrMetLeuLeuThrgluArgGlyLeuProArgArgArg1 228
   :::: |||::: |||||::: |||::: |||::: |||::: |||:::
2039 TCCTGCAATTCCTCTCTCTGTCACAAAGAGGCCCTGCCATGAACCCAA 2088
228 upheValMetGlnValLysValGlyHisHisThrAlaGluGlyValGlyT 245
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2089 GTTCCAATACTGTGTGAGTGGAGAGCCCAACTTCCCCAGTGTGAGTG 2138
245 hr...AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 260
   :: :|||: |||: |||: |||: |||: |||: |||: |||: |||:
2139 CTCCCAGCAAGAAAGTGGCAAGACAGATGCCGCCAGAGGCCATGAAG 2188
261 IleLeuGlyPheLysVal.....Progl 268
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2189 GCCCTGCATGGGAGGAGCCCAACTCCATGGCTTCTGATTAACCAAGCTGA 2238
268 nalaglnProalalysProalalaleuLysSerGluGluLysThrProValL 285
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2239 AGGTATGATCTCAGAGTCACTTGATTAACCTGGAATCCATGATGCCCAACA 2288
285 yslYsProglYaspGlyArgLysValThrPheGluProSerPro... 300
   || :||| :||| :||| :||| :||| :||| :||| :||| :|||
2289 AGGTCAGGAAGATGGGAGCTCGTAGATACCTGAACACCAACCCCTGTG 2338
300 ..... 300
2339 GGTGGCCTTTGGAGTACGCCCGCTCCCATGGCTTGTGCTGTAATTCAA 2388
301 ...GlyAspGluasnGlyThrSerAsnLysaspGluGluPheArgMetP 316
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
2389 GTTGGTCGACCAAGTCCGACCTCTCACCAG.....C 2420
316 roTyrLeuSerHisGlnGlnLeuProAlagly..IleLeuProMetVal 331
   || :||| :||| :||| :||| :||| :||| :||| :||| :|||
2421 CCAAGTTCGTTTACCAGCAAAAGTGGGGTGGCTGTTCCACAGCCGTC 2470
332 .....ProgluValAlaGlnAla..... 337
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
2471 TGGCGACACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCTGCT 2520
338 .....ValGlyValSerGlnGlnHisHisThrLysAspPheThrA 351
   ::||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2521 CCGTGTCTTGTATGGGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCTG 2570
351 rgAlaAlaProasnProalalalysAlaThrValThrAlaMetIleAlaArg 367
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
2571 AGGTAACCCCA.....GTGACAGGGGCCAGTCTCAGAAGA 2605
368 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSe 384
   ::||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2606 ACTATGCTCTCTCTCAAGGTCCCCAGAACACAG..... 2641
384 rasnIleSerSerGlyHisValProHisGlyProArgThrArgPro... 399
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
2642 .....CCAAAGACACTCCTCTCTCA 2660
400 .....SerGluGlnLeuTyrTyrLeuSerArgAlaGlnGly 411
   ::||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2661 CTGGCAGCAGCCTTCATGACCATGAGCATGCTGAGC..... 2698
412 pheGlnValGluTyrLysaspPheProLysasnAsnLysasnGluCysVa 428
   ::||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2699 .....CACCGTGTCTT 2709
428 lserLeuIleasnCysSerSerGlnProProLeuValSerHisGlyIleG 445
   :: :|||: |||: |||: |||: |||: |||: |||: |||: |||:
2710 CAACACTCTGACTAAGCTTCCAGCCCTCTTGTGCGGCCGCAAGATT. 2758
445 lylYsaspValGluSerCysHisaspMetAlaAlaLeuAsnIleLeuLys 461
   ::|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
2759 .....CTGGCCGCCATC..ATTATGAAA 2779

```

```

462 leuLeuSerGluLeuaspGlnGlnInsThrGluMetProArgThrGlyAs 478
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2780 AAAGACTCTGAG..GACATGGGTGTCTGCTCAGCTTGGGAACAGGAA 2826
478 nglyProValSerAlaCysGlyArgCys 487
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2827 T.....CGCTGT 2833

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-457-459-1
seq_documentation_block:
; Sequence 1, Application US/08457459
; Patent No. 5677428
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,459
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49CUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 155..3832
; US-08-457-459-1

alignment_scores:
Quality: 171.00 Length: 526
Ratio: 0.710 Gaps: 23
Percent Similarity: 45.817 Percent Identity: 21.483

alignment_block:
US-09-316-048-11 x US-08-457-459-1 ..
Align seg 1/1 to: US-08-457-459-1 from: 1 to: 6671
55 LysMetArgProProValLysHisAspAlaProAlaArgAla..... 68
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
1445 AGACTGAACCACTGTCATTACATGGCCCCCTCAAAGACAGGGTATGT 1494
69 .....LeuA 70
1495 TGACCTTGAAAATGGCCAGTGGGCCACAGATGACATCCAGATGACTTGA 1544
70 rgThrLeuGlnSerGluProLeuProGluArg....LeuGluValAsn 84
::::::::::|||:::||||:::||||:::
1545 ATAGTATCCGGCAGACACAGGTGAGTTTCGAGCCATCATGGAGATGCC 1594
85 Gly.....ArgGluAlaGln 89
:::||||:::
1595 TCCTTCTACAGTCATGCTTGCACAGGTGTCACCCCTACAAGAACTGAC 1644
89 uGluGluAsnLeuAsnLysSerGluIleSerGlnValPheGluIleAlaL 106
|||:::||||:::||||:::||||
1645 AGAGTGGCCAGCTG...AAGAACCCTACAGCGGCTGTAGATATGCCC 1691
106 euLysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 122
:::||||:::||||:::||||
1692 AGTTGCTAGTCAAACTGTGAGTTCACATGATAGACAGAGTGGACCA 1741
123 ProHisMetLysAsnPheValThrArgValSerVal..GlyGluPheVal 138
|||||:::||||:::||||:::||||
1742 CCCCATGAACCTCGATTAAATTCAGGTTGTCATCAATGCGCCGAGATT 1791
138 IGlyGluGlyGlyLysSerLysLysIleSerLysLysAsnAlaAla. 154
:::||||:::||||:::||||:::||||
1792 TCCCCCAGCTGAAGCTGGAAGCAAGAAAGTGGCCAAGCAGATGCACCTA 1841
155 .....ArgAlaValLeuGlnLeuArgArgLeuProProLeu 167
:::||||:::
1842 TGAAGCCATGACATTTCTGCTAGAGGAAGCCAAGCCAAGACAGTGA 1891
168 ProAlaValGluArgValLysProArgIleLysLysSerGlnProThr 184
:::||||:::||||:::||||
1892 AAATCAGAGAATCATCCCACTATTCCACAGAGAAAGAAATCAGAGAAC 1941
184 rCysLys.....ThrAlaProAspTyrGlyGlnGlyM 195
|:::||||:::||||:::||||
1942 TGCAGAGTCCAGACCCCAACCCCTTCAGCCACATCCTTTCTTGGA 1991
195 etAsnProIleSerArgLeuAlaGlnIleGlnAlaLysLysGluLys 211
:::||||:::||||:::||||
1992 AGAGCCCCGTACACACACTGCTTGAAGTATGAC...AAATTGGGGAAC 2038
212 GluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgL 228
:::||||:::||||:::||||
2039 TCCTGCGAATTCGCTCTCTGTCCAAAGAGGCCCTGCCCATGAACCCAA 2088
228 upheValMetGlnValLysValGlyHisHisThrAlaGluGlyValGlyT 245
:::||||:::||||:::||||
2089 GTTCCAATACTGTGTGAGTGGAGGCCCAACTTTCCCCAGTGTGAGTG 2138
245 hr..AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 260
:::||||:::||||:::||||
2139 CTCCAGCAAGAAAGTGCAGAAAGCAGATGGCCGACAGAGAACCATGAG 2188
261 IleLeuGlyPheLysVal.....ProGln 268
|||:::||||
2189 GCCCTGCATGGGAGCGACCAACTCCATGGCTTCTGATAACACAGCCTGA 2238
268 naIaGlnProAlaLysProAlaLeuLysSerGluGluLysThrProValL 285
::::::::::||||:::||||
2239 AGGTATGATCTCAGAGTCACTTGATTAAGTGAATCCATGATGCCCAACA 2288
285 yslLysProGlyAspGlyArgLysValThrPhePheGluProSerPro... 300
|||:::||||:::||||
2289 AGGTACAGGAAGATTGGCAGCTGTGAGATACCTGAACACCAACCTGTGT 2338
300 ..... 300
2339 GGTGGCCTTTGGAGTACGCCCGCTCCCATGGCTTTGCTGTAATTCAA 2388
```

```
301 ....GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetP 316
|||:::||||:::
2389 GTTGTGCAGACAGTCCGACCTCTCAGAG.....C 2420
316 rGlyrLeuSerHisGlnGlnLeuProAlaGly...IleLeuPrometVal 331
|||:::||||:::||||
2421 CCAAGTCTGTTTACCACGAAGAAAGTTGGGGGTCTGTTCCAGCCGTC 2470
332 .....ProGluValAlaGlnAla..... 337
|||:::||||
2471 TGGCAGACACAGCAAGAGCAAGGCAAGCAGGACAGATGCGGCTCT 2520
338 .....ValGlyValSerGlnGlyHisHisThrLysAspPheThrA 351
:::||||:::||||
2521 CCGTGTCTTGATTGGGAGAGACGAGAAGGACAGATGGTTTCACAG 2570
351 rGluAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 367
:::||||:::||||
2571 AGGTAAACCCA.....GTGACAGGGGCCAGTCTCAGAGA 2605
368 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSe 384
:::||||:::||||
2606 ACTATGCTCTCCTCTCAGAGTCCCAAGACAGACAG..... 2641
384 rAsnIleSerSerGlyHisValProHisGlyProArgThrArgPro.... 399
|||:::||||
2642 .....CCAAGACACTCCCTCTCA 2660
400 .....SerGluGlnLeuTyrTyrLeuSerArgAlaGlnGly 411
:::||||:::
2661 CTGGCAGCACCTTCCATGACAGATAGCATGCTGAGC..... 2698
412 PheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluCysVa 428
.....CACC GTGCTT 2709
2699 ..... 2709
428 IserLeuIleAsnCysSerSerGlnProProLeuValSerHisGlyIleG 445
:::||||:::||||
2710 CAACACTCTGACTAACAGCTTCCAGCCCTCTGCTCGGCCGCAAGATT. 2758
445 IyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 461
:::||||:::||||
2759 .....CTGGCCGCATC...ATTATGAAA 2779
462 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 478
|||:::||||
2780 AAAGACTCTGAG...GACATGGGTGTGCTGCTCAGCTTGGGAACAGGAA 2826
478 nGlyProValSerAlaCysGlyArgCys 487
|||:::
2827 T.....CGCTGT 2833
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-555-678-1

seq_documentation_block:

; Sequence 1, Application US/08555678

; Patent No. 5763174

; GENERAL INFORMATION:

; APPLICANT: Nishikura, Kazuko

; TITLE OF INVENTION: RNA Editing Enzyme and Methods

; NUMBER OF SEQUENCES: 67

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House

; CITY: Spring House

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,678
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,459
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49DUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 155..3832
; US-08-555-678-1
```

```

alignment_scores:
    Quality: 171.00      Length: 526
    Ratio: 0.710        Gaps: 23
    Percent Similarity: 45.817    Percent Identity: 21.483
```

alignment_block:
US-09-316-048-11 x US-08-555-678-1 ..

Align seg 1/1 to: US-08-555-678-1 from: 1 to: 6671

```

55 LysMetArgProProValLysHisAspAlaProAlaArgAla..... 68
   :::::::::::::::::::: ::::::::::::::::::::
1445 AGACTGAACCACTGTTTCATTACATGCCCCCTCAAAAGCAGGGTATGT 1494
69 .....LeuA 70
   ::::::::::::::::::::
1495 TGACCTTGAAATGCGCAGTGGCCACAGATGACATCCCATGACTTGA 1544
70 rGThrLeuGlnSerGluProLeuProGluArg.....LeuGluValAsn 84
   ::::::::::::::::::::
1545 ATAGTATCCGCGCAGCACACAGGTGAGTTTCAGCCATCATGAGATGCC 1594
85 Gly.....ArgGluAlaG1 89
   ::::::::::::::::::::
1595 TCCTTCTACAGTCATGGCTTGCACGCTGTCACCCCTACAGAAACTGAC 1644
89 uGluGluAsnLeuAsnLysSerGluIleSerGluValPheGluIleAla 106
   ::::::::::::::::::::
1645 AGAGTGCACAGTG...AAGAACCCCATCAGCGGCTGTAGAAATATGCC 1691
106 euLysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 122
1692 AGTTCGCTAGTCAAACTGTGAGTTCAACATGATAGAGCAGAGTGACCA 1741
123 ProHisMetLysAsnPheValThrArgValSerVal...GlyGluPheVa 138
   ::::::::::::::::::::
1742 CCCCATGAACCTCGATTAAATTCACAGGTTGTCATCAATGCGCCGAGAG 1791
```

```

138 IGIyGluGluGlyLysSerLysLysIleSerLysLysAsnAlaAla. 154
   ::::::::::::::::::::
1792 TCCCCCAGCTGAGCTGGAAGCAAGAAAGTGCCCAAGCAGAGATGCAGCTA 1841
155 .....ArgAlaValLeuGluGlnLeuArgArgLeuProLeu 167
   ::::::::::::::::::::
1842 TGAAGCCATGACAATTCGTCTAGAGAACCAAGCAAGCAGAGAGTGGA 1891
168 ProAlaValGluArgValLysProArgIleLysLysLysSerGlnProTh 184
   ::::::::::::::::::::
1892 AAATCAGAGAATCATCCCACTATTCACAGAGAAGAAGATCAGAGAAGAC 1941
184 rCysLys.....ThrAlaProAspTyrGlyGlnGlyM 195
   ::::::::::::::::::::
1942 TGCAGAGTCCCAAGACCCCAACCCCTTCAGCCACATCCTTCTTCTGGGA 1991
195 eLAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLys 211
   ::::::::::::::::::::
1992 AGAGCCCCGTCACACACTGCTTGAGTGTATGCAC...AAATTTGGGGAAC 2038
212 GluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgG1 228
   ::::::::::::::::::::
2039 TCCTGCAATTCCTGCTCTGTCCTGTCAAAGAGGCCCTGCCCATGAACCCA 2088
228 uPheValMetGlnValLysValGlyHisHisThrAlaGluGlyValGlyT 245
   ::::::::::::::::::::
2089 GTTCCAATCTGTGTTCAGTGGAGCCCAAACTTTCCCAAGTGTGAGTG 2138
245 hr...AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 260
   ::::::::::::::::::::
2139 CTCCAGACAAGAAAGTGCAAAAGCAGATGGCCGACAGAGAAGCATGGAAG 2188
261 IleLeuGlyPheLysVal.....ProG1 268
   ::::::::::::::::::::
2189 GCCCTGCATGGGAGGCGACCACTCCATGGCTTCTGATAACCAAGCCTGA 2238
268 nAlaGlnProAlaLysProAlaLeuLysSerGluGluLysThrProValL 285
   ::::::::::::::::::::
2239 AGGTATGATCTCAGAGTCACTTGATTAATTTGGAATTCATGATGCCCAACA 2288
285 yLysProGluAspGlyArgLysValThrPhePheGluProSerPro... 300
   ::::::::::::::::::::
2289 AGGTCAAGAAAGTGGCGAGCTCGTAGATACCTGAAACACCAACCTGTG 2338
300 ..... 300
2339 GGTGGCCTTTTGAGTACGCCCGCTCCCATGCGCTTGTGCTGAATTCAA 2388
301 ...GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetP 316
   ::::::::::::::::::::
2389 GTTGGTGCACAGTCCCGACCTCCTCACAGAG.....C 2420
316 rOTyrLeuSerHisGlnGlnLeuProAlaGly...IleLeuProMetVal 331
   ::::::::::::::::::::
2421 CCAAGTTCGTTTACCAAGCAAAAGTTGGGGTCCGCTGTTCCCAAGCCGTC 2470
332 .....ProGluValAlaGlnAla..... 337
   ::::::::::::::::::::
2471 TCGGCACACAGCAAGCAAGGCAAGCAAGCAAGCAAGCAGATGCGGCTCT 2520
338 .....ValGlyValSerGlnGlyHisHisThrLysAspPheThrA 351
   ::::::::::::::::::::
2521 CCGTGTCTTGATTTGGGAGAGAGAGAGAGAGAGAGCATGGGTTTCACAG 2570
351 rgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 367
   ::::::::::::::::::::
2571 AGTAACCCCA.....GTGACAGGGGCCAGTCTCAGAAGA 2605
368 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSe 384
   ::::::::::::::::::::
2606 ACTATGCTCTCTCTCAAGGTCCCAAGACACAGAG..... 2641
384 rAsnIleSerSerGlyHisValProHisGlyProArgThrArgPro... 399
```

```

2642 .....SerGlnLeuTyrTyrLeuSerArgAlaGlnGly 411
400 .....SerGlnLeuTyrTyrLeuSerArgAlaGlnGly 411
2661 CTGGCAGCACCTTCCATGACCAGATAGCCATGCTGAGC..... 2698
412 PheGlnValGluTyrLysAspPheProLysAsnLysAsnGluCysVal 428
2699 .....CACCAGTGTCTT 2709
428 LserLeuIleAsnGlySerSerGlnProProLeuValSerHisGlyIleG 445
2710 CAACACTCTGACTAACAGCTTCCAGCCCTCTTGTCTCGGCCGCAAGATT. 2758
445 LylAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 461
2759 .....CTGGCCGCCATC...ATTATGAAA 2779
462 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 478
2780 AAAGACTCTGAG...GACATGGGTGTCTGCTGACGCTTGGGACAGAGGAA 2826
478 nglyProValSerAlaCysGlyArgCys 487
2827 T.....CGCTGT 2833

```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02275-1

seq_documentation_block:

; Sequence 1, Application PC/TUS9502275

; GENERAL INFORMATION:

; APPLICANT: Wistar Institute of Anatomy & Biology

; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Cntr, P.O. Box 457

; CITY: Spring House

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/02275

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/280,443

; FILING DATE: 25-JUL-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/197,794

; FILING DATE: 17-FEB-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: WST49BPCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9206

; TELEFAX: 215-540-5818

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6671 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: cDNA

; FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 155..3832
PCT-US95-02275-1

```

```

alignment_scores:
    Quality: 171.00      Length: 526
    Ratio: 0.710        Gaps: 23
    Percent Similarity: 45.817    Percent Identity: 21.483

```

alignment_block:

US-09-316-048-11 x PCT-US95-02275-1 ..

Align seg 1/1 to: PCT-US95-02275-1 from: 1 to: 6671

```

55 LysMetArgProProValLysHisAspAlaProAlaArgAla..... 68
1445 AGACTGAAACACCACTGTTCAATTACAAATGGCCCTCAAAAGCAGGATATGT 1494
69 .....LeuA 70
1495 TGACTTTGAAAAATGGCCAGTGGCCACAGATGACATCCAGATGACTTGA 1544
70 rgtThrLeuGlnSerGluProLeuProGluArg.....LeuGluValAsn 84
1545 ATAGTATCCCGCGCAGCAGCAGGTGAGTTTCGAGCCATCATGAGATGCC 1594
85 Gly.....ArgGluAlaG1 89
1595 TCCTTCTACAGTCATGCTTGGCCACGGTGTTCACCCCTACAAGAACTGAC 1644
89 uGluGluAsnLeuAsnLysSerGlnIleSerGlnValPheGluIleAla1 106
1645 AGAGTGCAGCTG...AAGAACCCCATCAGCGGCTGTAGAATATGCC 1691
106 eulysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 122
1692 AGTTCGCTAGTCAAACTGTGAGTTCAACATGATAGACAGAGTGACCA 1741
123 ProHisMetLysAsnPheValThrArgValSerVal...GlyGluPheVal 138
1742 CCCCATGAACCTCGATTAAATTCACAGTTGTCAATCATGCGCCGAGAGTT 1791
138 lGlyGluGlnGluGlyLysSerLysLysIleSerLysLysAsnAlaAla. 154
1792 TCCCCAGCTGAAGCTGGAAGCAAGAAAGTGCCAGCAGATGCAGCTA 1841
155 .....ArgAlaValLeuGlnGlnLeuArgArgLeuProProLeu 167
1842 TGAAGCCATGACAAATTCTGCTAGAGGAAGCCAAAGCCAAAGACAGAGTGA 1891
168 ProAlaValGluArgValLysProArgIleLysLysLysSerGlnProTh 184
1892 AAATCAGAGAATCATCTCCACTATTCCACAGAGAAAGAAATCAGAGAAGAC 1941
184 rCysLys.....ThrAlaProAspTyrGlyGlnGlyM 195
1942 TGCAGAGTCCAGACCCCAACCCCTTCAGCCACATCTTCTTTCTGGGA 1991
195 eLAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLys 211
1992 AGAGCCCGTCACACACACTGTGAGTGTATGCAC...AAATTGGGAAC 2038
212 GluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgG1 228
2039 TCCTGCGAATTCCTCTCTCTGCAAGAAAGGCCCTGCCATGAACCCAA 2088
228 uPheValMetGlnValLysValGlyHisIleThrAlaGluGlyValGlyT 245
2089 GTTCCAATACTGTGTGACAGTGGAGCCCAAACTTTCCCAAGTGTGAGTG 2138
245 hr..AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 260
; .....

```

```

2139 CTCCAGCAAGAAAGTGGCAAGCAGATGGCCGACAGAGGAAGCCATGAAG 2188
261 ILeuGlyPheLysVal.....ProG1 268
      |||
2189 GCCCTGCATGGGAGCGACCAACTCCATGCTTCTGTATACCAAGCCTGA 2238
      |||
268 nAlaGlnProAlaLysProAlaLeuLysSerGluGluLysThrProVal 285
      |||
2239 AGGTATGATCTCAGAGTCACTTGATTAATGGAATCCATGATGCCCAACA 2288
      |||
285 yLysProGlyAspGlyArgLysValThrPhePheGluProSerPro... 300
      |||
2289 AGGTCAGGAAGATTGGCGAGCTCGTAGATACCTGAACACCAACCCCTGTG 2338
      |||
300 ..... 300
2339 GGTGGCCTTTGGAGTACGCCCGCTCCATGGCTTGTGCTGTAATTCAA 2388
301 ...GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetP 316
      |||
2389 GTTGTGCACCACTCCGACCTCCTCAGAG.....C 2420
316 roTyLLeuSerHisGlnGlnLeuProAlaGly...IleLeuProMetVal 331
      |||
2421 CCAAGTTGTTTACCAAGCAAAAGTTGGGGTCCGTTCCAGCCGCTC 2470
332 .....ProGluValAlaGlnAla..... 337
      |||
2471 TGGCGACACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2520
338 .....ValGlyValSerGlnGlyHisHisThrLysAspPheThra 351
      |||
2521 CCGTGTCTTGATTGGGAGAGAACGAGAGCAAGCAAGCAAGCAAGCAAGCA 2570
351 rGAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 367
      |||
2571 AGGTAACCCCA.....GTGACAGGGGCCAGTCTCAGAAGA 2605
368 GluLeuLeuTyrglyGlyThrSerProThrAlaGluThrIleLeuLysSe 384
      |||
2606 ACTATGCTCTCTCTCAAGTCCCGCAAGACACAG..... 2641
384 rAsnIleSerSerGlyHisValProHisGlyProArgThrArgPro.... 399
      |||
2642 .....CCAAAGACACTCCCTCTCA 2660
400 .....SerGluGlnLeuTyrglyLeuSerArgAlaGlnGly 411
      |||
2661 CTGGCAGCACCCTTCATGACAGATAGCCATGCTGAGC..... 2698
412 PheGlnValGluTyrglyAspPheProLysAsnAsnLysAsnGluCysVa 428
      |||
2699 .....CACCGGTGCTT 2709
428 IserLeuIleAsnCysserSerGlnProProLeuValSerHisGlyIleg 445
      |||
2710 CAACACTCTGACTAACAGCTTCCAGCCCTCTGCTCGGCCGCAAGATT. 2758
445 lYlYAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 461
      |||
2759 .....CTGGCCGCCATC..ATTATGAAA 2779
462 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 478
      |||
2780 AAAGACTCTGAG...GACATGGGTGCTGCTCAGCTTGGGACAGGGAA 2826
478 nGlyProValSerAlaCysGlyArgCys 487
      |||
2827 T.....CGCTGT 2833

```

seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-931-999-4

seq_documentation_block:

```

; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: UT0007
; US-08-931-999-4

alignment_scores:
      Quality: 129.00      Length: 379
      Ratio: 0.754      Gaps: 12
Percent Similarity: 45.119      Percent Identity: 20.580

alignment_block:
US-09-316-048-11 x US-08-931-999-4 ..
Align seg 1/1 to: US-08-931-999-4 from: 1 to: 6755

84 AsnGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluIleSerG1 100
      |||
126 AACACGCGACACCGCGCGAGGAGGAGAGCAAGCAAGACACAGAAAGA 175
100 nValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluVal. 116
      |||
176 AGAACGAGAACAAAAACAGAAAAAGAACCAAAAGAAAGAAAGAAACGA 225
117 ....AlaArgGluSerGlyProProHisMetLysAsnPheValThrArg 131
      |||
226 ACCCAGAAAAAGAGAGAGCAAGAGAGAAAGAACACCAAAAGCGAGA 275
132 ValSerValGlyGluPheValAlGlyGluGlyGlyLysSerLysI1 148
      |||
276 AAAGCAAGGCGGAGACAAACGGAACCAAGCAAAACCAAAACCAAGCG 325

```



```

156 AlavalLeuGluInLeuArgLeuProProLeuProAlaValGluAr 172
      ::::::::::::::
394 TTAGCTGTGAGATACTTAAT..... 414
172 gValLysProArgLleLysLysSerGlnProThrCysLysThrAlaP 189
      ::::::::::::::
415 .....AAGAAAGAAGAGCAGTTAGTCCTTTATTATTGACACAA 454
189 roAspTyrGlyGlnGlyMetAsnPro.....IleSer 199
      ::::::::::::::
455 CGAATTCTTCAGAGAGATTATCCATGGGGAATTACATAGCCCTTATCAAT 504
200 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrMe 216
      ::::::::::::::
505 AGAATTGCCAG.....AAGAAAGACTAACTGTAAATTATGA 542
216 tLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetGlnV 233
      ::::::::::::::
543 ACAGTGTGCATCG...GGGTCGATGGCCAGAGATTTCATTATTAAT 589
233 alLysValGlyHis...HisThrAlaGluGlyValGlyThrAsnLysLys 248
      ::::::::::::::
590 GCAAAATGGGACAGAAACAATATAGTATTGGTACAGGTTCTACTAAACAG 639
249 ValAlaLysArgAsnAlaAlaGlu..AsnMetLeuGluIleLeuGlyPn 264
      ::::::::::::::
640 GAAGCAAAACAATTGGCCGCTAAACTTGCAATATCTTCAGATATTA.... 684
264 elYsValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluL 281
      ::::::::::::::
685 .....TCAGAAG 691
281 ysThrProValLysLysPro.....GlyAspGlyArgLysValThrPhe 295
      ::::::::::::::
692 AAACCTCAGTGAATCTGACCTACCTGCTCCTGCTTCTTTGCTACTACG 741
296 pheGlu..... 297
      ::::::::::::::
742 TGTGAGTCCCAAGCACTCTTAGTGACACGACACTCGCTTGTGAATC 791
298 .ProSerProGlyAsp.....GluAsnG 305
      ::::::::::::::
792 ATCATCTGAAGGTACTTCTCAGCAGATACATCAGATAAATTCTAACA 841
305 lyThrSerAsnLysAspGluGluPheArgMetProTyrLeuSerHisGln 321
      ::::::::::::::
842 GTGACAGTTTAAACAGTCTCTGCTTATGAATGGTCTCAGAAATTAAT 891
322 Gln.....LeuProAlaGlyIleLeuProMetVa 331
      ::::::::::::::
892 CAAGGAAGCAAAAAAGATCTTTGGCACCAGATTGTGACCTTCCGTGACAT 941
331 lProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysA 348
      ::::::::::::::
942 GAAAGAAACAAGTATCTGTGACACAAGAGTTTGGCATGATTTTAAAG 991
348 sp.....PheThrArgAlaAlaProAsnPro 356
      ::::::::::::::
992 AAATAGAATTAAATGGCTCAGGTGGATTTGGCCAAAGTTTCAAAAGCAAAA 1041
357 AlalysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyG1 373
      ::::::::::::::
1042 CACAGAAATTGACGGAAGACTTACGTTATTAACGTTAAATAT.... 1086
373 yThrSerProThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyH 390
      ::::::::::::::
1087 .AATAACGAGAAGCGGCGGTGAAGTAAAGCATTTGCCAAACCTGATC 1135
390 lValProHisGlyProArgThrArgProSerGluGlnLeuTyrTyrLeu 406
      ::::::::::::::
1136 ATGTA.....AATATTGTTCACTACAT 1158

```

```

407 SerArgAlaGlnGlyPhe.....GlnVa 414
      ::::::::::::::
1159 GCCTGTGGGATGGATTTGATTATGATCCTGAGACCAGTGATGATTCTCT 1208
414 lGluTyrLysAspPhe...ProLysAsnAsnLysAsnGluCysValSerL 430
      ::::::::::::::
1209 TGAGAGCAGTGAATTATGATCTGTGAGAACAGCAAAATAGTTCAAGGTCAA 1258
430 euIleAsnCys 433
      ::::::::::::::
1259 AGACTAAGTGC 1269
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:us-08-434-998-8
seq_documentation_block:
; Sequence 8, Application US/08434998
; Patent No. 5866781
; GENERAL INFORMATION:
; APPLICANT: Silverman, Robert H.
; APPLICANT: Sengupta, Dibyendu N.
; TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
; TITLE OF INVENTION: Cells and Methods
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
; ADDRESS: Russell
; STREET: 200 E. Broward Boulevard
; CITY: Fort Lauderdale
; STATE: Florida
; COUNTRY: USA
; ZIP: 33301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,998
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,973
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Manso, Peter J.
; REGISTRATION NUMBER: 32,264
; REFERENCE/DOCKET NUMBER: CL11363-16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 305/527/2498
; TELEFAX: 305/764/4996
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2562 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-434-998-8
alignment_scores:
Quality: 122.50 Length: 387
Ratio: 0.677 Gaps: 17
Percent Similarity: 46.770 Percent Identity: 22.481
alignment_block:
US-09-316-048-11 x US-08-434-998-8 ..
Align seg 1/1 to: US-08-434-998-8 from: 1 to: 2562
107 LysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProPr 123
      ::::::::::::::
244 AAGCAGGAGTAGTACTTAATATCAAGAACTGCCTAATTCAAGACCTCC 293

```



```

123 OHISmetLysAsnPhenValThrArgValSerVal...GlyGluPheValG 139
124 ACATGATAGAGGTTTACATTTCAGTTATATAGATGGAAGAGATTTTC 343
125 1YGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUG 155
126 CAGAAGGTGAAGGTAGATCAAGAAGCAAAAAATGCCGACGCCAAA 393
127 A1AValLeuGluGluLeuArgArgLeuProProLeuProAlaValGluAr 172
128 TTACCTGTGAGATCTTAAT... 414
129 gValLysProArgIleLysLysLysSerGlnProThrCysLysThrAlaP 189
130 AAGAAAAGAGAGGAGGAGTTCCTTATTTATTTGACAAACAA 454
131 roAspTyrGlyGlnGlyMetAsnPro...IleSer 199
132 CGAATTCCTTCAGAAGATATCCATGGGAATTCATAGCCTTATCAAT 504
133 ArgLeuAlaGlnIleGlnAlaLysLysGluLysGluProGluTyrMe 216
134 AGAATTGCCAG...AAGAAAGACTACTGTAATATATGA 542
135 tLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGlnV 233
136 ACAGTGTGCATCG..GGGTGCATGGCCAGAGAGATTTCTATTAAT 589
137 allysValGlyHis..HisThrAlaGluGlyValGlyThrAsnLysLys 248
138 GCAAAATGGGACAGAAAGATATAGTATTGGTACAGTTCTACTAAACAG 639
139 ValAlaLysArgAsnAlaAlaGlu...AsnMetLeuGluIleLeuGlyPh 264
140 GAAGCAAAACAATTTGGCCGCTTAACCTTCATATCTTCAGATATTA.... 684
141 eLysValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluL 281
142 TGTGAGTCCCAAAGCAACTCTTTAGTGACACAGACACTGCTTCTGAATC 791
143 ProSerProGlyAsp...GluAsnG 305
144 ATCATCTGAAGGTGACTTCTCAGCAGATACATCAGAGATAAATTTCTAAC 841
145 1YThrSerAsnLysAspGluGluPheArgMetProTyrLeuSerHisGln 321
146 GTGACAGTTTAACAGTTCCTGCTTATGATGCTCAGAAATAT 891
147 Gln.....LeuProAlaGlyIleLeuProMetVa 331
148 CAAAGGAAAGCAAAAGATCTTTGGCACCAGATTGACCTTCTGCACAT 941
149 1ProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysA 348
150 GAAAGAAACAAAGTATCTGACCAAGAGGTTTGGCATGATTTTAAAG 991
151 SP.....PheThrArgAlaAlaProAsnPro 356
152 AAATAGAAATTAATTGGCTCAGGTGATTTGGCCAAGTTTCAAAGCAAAA 1041
153 AlAlysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyG 373
154 CACAGAAATTGACGGAAGAGACTTACGTTATTAAACGTGTTAAATAT.... 1086

```

```

373 YThrSerProThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyH 390
374 AATAACGAGAAGCGCGAGCGTGAAGTAAAGCAATTCGCAAAACTTGATC 1135
375 1sValProHisGlyProArgThrArgProSerGlnLeuTyrTyrLeu 406
376 ATGTA.....AATATGTTCACTACAAAT 1158
377 SerArgAlaGlnGlyPhe.....GlnVa 414
378 GGCTGTGGATGATTTGATATGATCCTGAGACCAAGTATGATTCCTCT 1208
379 1GluTyrLysAspPhe..ProLysAsnAsnLysAsnGluCysValSerL 430
380 TGAGAGCAGCATGATATGATCCTGAGACCAAGCAAAATAGTTCAAGTCAA 1258
381 euLLeAsnCys 433
382 AGACTAGTGC 1269

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-487-797-8

seq_documentation_block:

```

: Sequence 8, Application US/08487797
: Patent No. 5866787
: GENERAL INFORMATION:
: APPLICANT: Silverman, Robert H.
: TITLE OF INVENTION: Transgenic Plants Co-Expressing A
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Holland & Knight
: STREET: One E. Broward Boulevard, #1300
: CITY: Fort Lauderdale
: STATE: Florida
: COUNTRY: USA
: ZIP: 33301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,797
: FILING DATE: 07-JUNE-1995
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Manso, Peter J.
: REGISTRATION NUMBER: 32,264
: REFERENCE/DOCKET NUMBER: CL11363-16(C)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 305/468-7811
: TELEFAX: 305/463-2030
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2562 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-487-797-8

```

alignment_scores:

Quality:	Ratio:	Length:
122.50	0.677	387
46.770	Percent Similarity:	Gaps: 17
	Percent Identity:	22.481

alignment_block:

US-09-316-048-11 x US-08-487-797-8 ..

Align seg 1/1 to: US-08-487-797-8 from: 1 to: 2562

```

107 LysArgAsnLeuProValAsnPhelValAlaArgGluSerGlyProPr 123
    |||:::|||||:::|||||:::|||||
244 AAGCAGGAGTAGTACTTAATATCAAGAACTGCCTAATTCAGACCTCC 293
    |||:::|||||:::|||||:::|||||
123 OHISmetLysAsnPhelValThrArgValSerVal..GlyGluPhelValG 139
    |||:::|||||:::|||||:::|||||
294 ACATGATAGGAGGTTTACATTTTCAAGTTATAATAGATGAGAGAAATTTC 343
    |||:::|||||:::|||||:::|||||
139 LysGluGlyGluGlyLysSerLysLysIleSerLysLysAsnAlaAlaArg 155
    |||:::|||||:::|||||:::|||||
344 CAGAAAGGTGAAGGTAGATCAAAAGAAAGAAATAATGCCGACAGCCAAA 393
    |||:::|||||:::|||||:::|||||
156 AlaValLeuGluGluLeuArgArgLeuProProLeuProAlaValGluAr 172
    |||:::|||||:::|||||:::|||||
394 TTAGCTGTGTGAGATACTTAAT..... 414
    |||:::|||||:::|||||:::|||||
172 gValLysProArgIleLysLysLysSerGluProThrCysLysThrAlaP 189
    |||:::|||||:::|||||:::|||||
415 .....AAGGAAAGAAAGGACGAGTTAGTCTTTATTATTGACACACAA 454
    |||:::|||||:::|||||:::|||||
189 roAspTYrGlyGlnGlyMetAsnPro.....IleSer 199
    |||:::|||||:::|||||:::|||||
455 CGAATTCTTCAGAAAGGATTTATCCATGGGGAATTACATAGCCCTTATCAAT 504
    |||:::|||||:::|||||:::|||||
200 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTYrMe 216
    |||:::|||||:::|||||:::|||||
505 AGAATTGCCAG.....AAGAAAAGACTAAGTGTAAATTATGA 542
    |||:::|||||:::|||||:::|||||
216 tLeuLeuThrGluArgGlyLeuProArgArgArgGluPhelValMetGlnV 233
    |||:::|||||:::|||||:::|||||
543 ACAGTGTGCATCG...GGGTCGCATGGGCCAGAGCATTTTCATTATATAAT 589
    |||:::|||||:::|||||:::|||||
233 allYsValGlyHis..HisThrAlaGluGlyValGlyThrAsnLysLys 248
    |||:::|||||:::|||||:::|||||
590 GCAAAATGGGACAGAAAGATATAGTATTTGTTACAGGTTCTACTAAACAG 639
    |||:::|||||:::|||||:::|||||
249 ValAlaLysArgAsnAlaAlaGlu..AsnMetLeuGluIleLeuGlyPh 264
    |||:::|||||:::|||||:::|||||
640 GAAGCAAAACAATTGGCCGCTAAACTTGCATATCTTCAGATATTA.... 684
    |||:::|||||:::|||||:::|||||
264 eLysValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluL 281
    |||:::|||||:::|||||:::|||||
685 .....TCAGAAAG 691
    |||:::|||||:::|||||:::|||||
281 ySThrProValLysLysPro....GlyAspGlyArgLysValThrPhe 295
    |||:::|||||:::|||||:::|||||
692 AAACCTCAGTGAATCTGACTACCTGTCCTGTGTTCTTTGCTACTACG 741
    |||:::|||||:::|||||:::|||||
296 pheGlu..... 297
    |||:::|||||:::|||||:::|||||
742 TGTGAGTCCCAAAGCAACTTTAGTGACCAGACACACTCGCTTGTGAATC 791
    |||:::|||||:::|||||:::|||||
298 .ProSerProGlyAsp.....GluAsnG 305
    |||:::|||||:::|||||:::|||||
792 ATCATCTGAAGTGACTTCTCAGCAGATACATCAGAGATAAATTTCTAACA 841
    |||:::|||||:::|||||:::|||||
305 LysThrSerAsnLysAspGluGluPheArgMetProTYrLeuSerHisGln 321
    |||:::|||||:::|||||:::|||||
842 GTGACAGTTTAACAGTTCTTCGTTGCTTATGAAATGCTCTCAGAAATAAT 891
    |||:::|||||:::|||||:::|||||
322 Gln.....LeuProAlaGlyIleLeuProMetVal 331
    |||:::|||||:::|||||:::|||||
892 CAAAGGAAGGCAAAAAGACTTTTGGCACCAGATTGACCTTCTGACAT 941
    |||:::|||||:::|||||:::|||||
331 lProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysA 348
    |||:::|||||:::|||||:::|||||
942 GAAAGAAACAAGATATCTGTGACAAAGAGTTTGGCATGATTTTAAAG 991
    |||:::|||||:::|||||:::|||||
348 sp.....PheThrArgAlaAlaProAsnPro 356
    |||:::|||||:::|||||:::|||||
992 AAATAGATTAATTGGCTCAGGTGATTTGGCCAAAGTTTTCAAAGCAAAA 1041
    |||:::|||||:::|||||:::|||||

```

```

357 AlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTYrGlyG1 373
    |||:::|||||:::|||||:::|||||
1042 CACAGAAATTGACGGAAGACTTACGTTATTAAAGCTGTTAAATAT..... 1086
    |||:::|||||:::|||||:::|||||
373 yThrSerProThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyH 390
    |||:::|||||:::|||||:::|||||
1087 .ATAACGAGAAGCGCGAGCGTGAAGTAAAGCATTTGGCAAAACTTGATC 1135
    |||:::|||||:::|||||:::|||||
390 lValProHisGlyProArgThrArgProSerGluGlnLeuTYrTYrLeu 406
    |||:::|||||:::|||||:::|||||
1136 ATGTA.....AATATGTTCTACTACAT 1158
    |||:::|||||:::|||||:::|||||
407 SerArgAlaGlnGlyPhe.....GlnVa 414
    |||:::|||||:::|||||:::|||||
1159 GGCTGTGGATGATTTGATTATGATCCTGAGACAGCATGATGATTTCTCT 1208
    |||:::|||||:::|||||:::|||||
414 lGluTYrLysAspPhe...ProLysAsnAsnLysAsnGluCysValSerL 430
    |||:::|||||:::|||||:::|||||
1209 TGAGAGCAGTGATTTATGATTCCTGAGAACAGCAAAATAGTTCAAGGTCAA 1258
    |||:::|||||:::|||||:::|||||
430 euleAsnCys 433
    |||:::|||||:::|||||:::|||||
1259 AGACTAACTGC 1269
    |||:::|||||:::|||||:::|||||
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02058-8
seq_documentation_block:
; Sequence 8, Application PC/TUS9502058
; GENERAL INFORMATION:
; APPLICANT: Silverman, Robert H.
; APPLICANT: Sengupta, Dibendu N.
; TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
; TITLE OF INVENTION: Cells and Methods
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
; ADDRESSEE: Russell
; STREET: 200 E. Broward Boulevard
; CITY: Fort Lauderdale
; STATE: Florida
; COUNTRY: USA
; ZIP: 33301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02058
; FILING DATE:
; *CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198, 973
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Manso, Peter J.
; REGISTRATION NUMBER: 32,264
; REFERENCE/DOCKET NUMBER: CL11363-16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 305/527/2498
; TELEFAX: 305/764/4996
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2562 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US95-02058-8

```

alignment_scores: 122.50 length: 387


```

; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HUMAN PKR GENE, FIGURE 5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 187..1836
; US-08-143-219-1

```

```

alignment_scores:
    Quality: 121.50      Length: 188
    Ratio: 1.191        Gaps: 8
    Percent Similarity: 54.255    Percent Identity: 27.128

```

```

alignment_block:
US-09-316-048-11 x US-08-143-219-1 ..

```

Align seg 1/1 to: US-08-143-219-1 from: 1 to: 2628

```

107 LysArgAsnLeuProValAsnPhenGluValAlaArgGluSerGlyProPr 123
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
244 AAGCAGGAGTAGTACTTAATATCAAGACTGCCCTAATTCAGGACCTCC 293
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
123 CHISMetLysAsnPhenValThrArgValSerVal..GlyGluPheValG 139
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
294 ACATGATAGGAGGTTTACATTTCAGTTATAATAGATGGAAGAGATTTC 343
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
139 IyGluGlyGlyGlySerLysLysLysLysLysLysLysLysLysLysLys 155
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
344 CAGAAAGTGAAGGTAGATCAAGAAGGAGCAAAATGCCGACGCCAA 393
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
156 AlavalLeuGluGluLeuArgArgLeuProProLeuProAlavalGluAr 172
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
394 TTAGCTGTTGAGATACTTAAT..... 414
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
172 gValLysProArgLysLysLysLysSerGlnProThrCysLysThrAlaP 189
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
415 .....AAGAAAGAAGAGCAGTTAGTCCTTATTATTGACACAA 454
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
189 roAspTyrGlyGlnGlyMetAsnPro.....IleSer 199
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
455 CGAATTCTTCAGAGAGATTATCCATGGGGAATTACATAGCCCTTATCAAT 504
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
200 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrMe 216
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
505 AGAATTGCCAG.....AAGAAAGACTAATCTGTAATTAATGA 542
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
216 tLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGlnV 233
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
543 ACAGTGTGCATCG...GGGGTCATGGCCAGAAAGATTTCATTATTAAT 589
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
233 alLysValGlyHis...HisThrAlaGluGlyValGlyThrAsnLysLys 248
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
590 GCAAAATGGGACAGAAAGAATATAGTATGTGTACAGGTTCTACTAACAG 639
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
249 ValAlaLysArgAsnAlaAlaGlu..AsnMetLeuGluIleLeuGlyPh 264
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
640 GAAGCAAAACAATGGCCGCTAACTGCAATATCTTCAGATATTA..... 684
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
264 elYsValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluL 281
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
685 .....TCAGAAAG 691
281 ySThrProValLys 285

```

```

::|||  |||||
692 AAACCTCAGTGAA 705

```

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-717-294-42

seq_documentation_block:

; Sequence 42, Application US/08717294

; Patent No. 6114148

; GENERAL INFORMATION:

; APPLICANT: SEED, BRIAN

; APPLICANT: HAAS, JURGEN

; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF

; NUMBER OF SEQUENCES: 110

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Clark & Elbing LLP

; STREET: 176 Federal Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/717,294

; FILING DATE: 20-SEP-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Elbing, Karen L

; REGISTRATION NUMBER: 35,238

; REFERENCE/DOCKET NUMBER: 00786/345001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-428-0200

; TELEFAX: 617-428-7045

; TELEX:

; INFORMATION FOR SEQ ID NO: 42:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4451 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-717-294-42

```

alignment_scores:
    Quality: 117.00      Length: 522
    Ratio: 0.539        Gaps: 30
    Percent Similarity: 41.571    Percent Identity: 22.031

```

alignment_block:

US-09-316-048-11 x US-08-717-294-42 ..

Align seg 1/1 to: US-08-717-294-42 from: 1 to: 4451

```

31 ProPheProValPro.....ProLeuLeuTyrGlnValG1 42
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
167 CCGCTTCCCCCGCGGTGCCCAAGAGCTTCCCTTCAACACACAG..... 211
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
42 uLeuSerValGlyGlyGlnGlnPheAsnGlyLysGlyLysMetArgProp 59
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
212 .....CGTGTGTACAAAGAAACCCCTGTTGAGTTACACGACAC 254
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
59 roValLysHis.....AspAlaProAla..... 66
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
255 CTGTCAACATTTGCCAAGCGCGCCGCCCTGATGGCCCTGCTGGGCC 304

```

```

67 .....ArgAlaLeuArgThrLeuGlnSerGluPro..... 76
   |||:::||||
305 CACCATCCAGGCCGAGGTGTACGACACCCGTGGTGTACCCCTGAAGAACA 354
77 .....LeuProGluArgLeuGluValAsnGlyArgGluAlaGluGlu 91
   |||:::||||
355 TGGCCAGCCACCCCGTCAGCCTGCACGCCGTGGCGGTGAGCTACTGGAAG 404
91 LuAsnLeuAsnLysSerGluIleSerGlnValPheGluIleAlaLeuLys 107
   :::::
405 GCCAGCGAGGGCGCGCA.....GTACGACGA 430
108 ArgAsnLeuProValAsnPheGluValAlaArgGlu..... 119
   :::::||||
431 CCAGACGTCCCA.....GCCGAGAAGAGAGACGACACAAG 465
120 .....SerGlyProProHisMetLysAsnPro..... 128
   ::||| |||:::
466 TGTTCCTCCGGGGGAGGCCACCTACGTGTGGCAGCTGCTTAAGAGAAC 515
129 .....ValThrArgValSerValGlyGluPhe 137
   |||:::
516 GGCCTATATGGCCAGCGACCCCTGTGCTGACCTACAGCTACCTGAGCCA 565
138 ValGlyGluGlyGlySerLysSerLysLysIleSerLysLysAsnAla.. 153
   ||| ||| |||:::
566 CGTGCACCTGTGTGAAGATCTGAACAGCGGCTGATCGCGCCTGCTGG 615
154 ....AlaArgAlaValLeuGlnLeuArgArgLeuProProLeu... 167
   ::|||::: |||:::
616 TGTGTCCGCGAGGCGAGCCTGCCCAAGAGAAAACCCAGACCCCTGCACAAG 665
168 .....ProAlaValGluArgValLysProArgIleLysLysSerGln 182
   ||| ||| |||:::
666 TTCATCTGCTGTTCGCCGTGTTCA..CGAGGGGAAGAGCTGGCACAG 712
183 ProThrCysLysThrAlaProAspTyrGlyGlnGlyMetAsnProIleSe 199
   ::||| |||
713 CGAGACTAAGAA..CAGCTGATGCAGAGCGCGCGCCGCGCAGCGCCC 759
199 rArgLeuAlaGlnIleGlnAlaLysLysGluLysGlu..... 212
   ||| ||| |||
760 GCGCCTGGCCCAAGATGCACACCGTTAACGGCTAGCTGAACCGCAGCCTG 809
213 .....ProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArg 227
   |||::: ||| ||| |||
810 CCGGCGCTGATCGGCTGCCACCGCAAGAGCGGTACTGGCACGTCATCGG 859
228 GluPheValMetGlnValLysValGlyHisThrAlaGluGlyValGlu 244
   ::||| ||| |||
860 CAT.....GGGCACCACCCC..... 874
244 yThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMet..... 258
   ::|||:::
875 .....TGAGGTGCACAGCATCTTCCCTGGAG 899
259 ..LeuGluIleLeuGlyPheLysValProGlnAlaGlnProAlaLys... 273
   ::::: |||::: |||
900 GGCACACCTTCTGCTGCGCAACACCGCCAGCGCAGCCTTGAGATCAG 949
273 ..... 273
950 CCCCATCACCTTCTGACTGCCAGACCCCTGCTGATGACCTAGGCCAGT 999
274 .ProAlaLeuLysSerGluGluLysThrProValLysLysProGly... 288
   ||| |||::: |||::: |||
1000 TCCGTGCTGTCTGCCACATCAGCAGCCAGCAGCAGCGCATGGAGGCT 1049
289 .....AspGlyArgLysValThrPhePheGluProSerProGlyAspGlu 303
   ::|||::: |||::: |||
1050 TACGTGAAGGTGGACAGCTGCCCGAGAGAGCCCGCAGCTGCGCATGAAGAA 1099
304 .....AsnGlyThrSerAsnLysAspGluGluPheArg..... 314

```

```

1100 CAACGAGAGGCCGAGACTACGACGACGACCTGACCGACAGCAGATGG 1149
   ::|||:::
315 .....MetProTyrLeuSer..... 319
1150 ATGTCTACGCTTCGACGACGACAACAGCCCGAGCTTCATCCAGATCCGC 1199
320 .....HisGlnGln... 322
   |||:::
1200 AGCGTGCCCAAGAGACCCCTAAGACCTGGGTGCACCTACATCGCCGCCGA 1249
323 .....LeuProAlaGlyIleLeuProMetValProGlu 333
   ||| |||:::
1250 GGAGGAGACTGGGACTACGCCCGCTAGTACTGGCCCCGACGACGCCA 1299
333 uValAlaGlnAlaVal..... 338
   ::|||
1300 GCTACAGAGCCAGTACCTGACACACGCGCCCGACGCGATCGCGCCAAG 1349
339 .....GlyValSerGlnGlyHisThrLysAspPheThrArgAla 352
   |||::: ||| |||:::
1350 TACAAGAGGTGCGCTTCATGGCTTACACCGACGAGACTTCAAGACCCG 1399
353 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 369
   ::||| |||:::
1400 CGAGGCCATCCAGACGAGCTCGGCATCTCTGGC.....CCCC 1437
369 euLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSerAsn 385
   ||| ||| |||:::
1438 TGTCTGACGGCGAGGTGGGCGACACCTGCTGATCATCTTCAAGAACCAG 1487
386 IleSerSerGlyHis.....ValProHisGly..ProArgThrArgPr 399
   |||::: ||| ||| |||
1488 GCCAGCAGGCCCTACACATGTACCCCGACGCGCATCAGCGACGTGCGCCC 1537
399 oSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGlu 416
   ||| ||| |||:::
1538 C.....CTGTACAGCGCGCGCTGCCCCAAGGGCGTGAAG..CACC 1575
416 yrlYsAspPhePro 420
   ||| ||| ||| |||
1576 TGAAGGACTTCCCC 1589

```

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-335-409-1

seq_documentation_block:

```

; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

```

alignment_scores:

Quality:	116.50	Length:	442
Ratio:	0.604	Gaps:	20
Percent Similarity:	43.665	Percent Identity:	20.814

alignment_block:
US-09-316-048-11 x US-09-335-409-1/rev ..

Align seg 1/1 to reverse of: US-09-335-409-1 from: 1 to: 68750

```

      8 HisSerArgMetGlnSerThrTyrSerTyrGlyMetArgGlyGlyAlaTyr 24
      51360 CACTTGAGGATGCCATGTCTCTCGTGA...AGCGTACCCGGCCGACGACC 51314
      24 rProProArgTyrPheTyrProPheProValProProLeuLeuTyrGlnV 41
      51313 TCCCCCGTCCGAGCCACACCTGTCTCCAGCCCT..... 51280
      41 alGluLeuSerValGlyGlyGlnPheAsnGlyLysGlyLysMetArg 57
      51279 .....GGAAGCAGGAGCACCATAAGTCAGGGCGTCTCGT 51247
      57 ..... 57
      51246 TGAGCAGCGGATAGAGCGCCTCCGACGACGAGCGCTCGACATCGCCGT 51197
      58 .....ProProValLysHisAspA 64
      51196 TGAATCTCCACGAGCTTCTCTCAACCGCTCCACCGCTGAGCTGTGGCC 51147
      64 laProAlaArgAla...LeuArgThrLeuGlnSerGluProLeuProGlu 79
      51146 GCTACACACGTGCCCGTGGCGTGGCGGACCCAGCTCTTACCTGCTCCT 51097
      80 ArgLeuGluValAsnGlyArgGluAlaGluGluAsnLeuAsnLysSe 96
      51096 CACGACTCGATACCTGGAATGAGGCCGCGGACCTTCTTCTGCTGAGG 51047
      96 rGluLeuSerGlnValPheGluLeuAlaLeuLysArgAsn..... 109
      51046 GCCACTTGACAGATGGGTCCGCTTCGAGGAGCAGCGGAGGGCTGCTC 50997
      110 .....LeuProValAsnPheGluValAla 117
      50996 GAGCGCCAGCTGCTCCAGCACGACGCTCGCCGTGCCATAGAGATCTACGC 50947
      118 ArgGluSerGlyProPro...HisMetLysAsnPheValThrArgValSe 133
      50946 CGGCGGCGAGCGCCATCTCTACATACGCGCGCTGGCAACACGCGCTTCG 50897
      133 rValGlyGluPheValGlyGlyGlyLysSerLysLysIleSerL 150
      50896 CCATGAGACCTATGTTCGAAAGT.....AGATAGCCT 50862
      150 ySLysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProPro 166
      50861 CTCGTCGCTCAGCGCTTGTCTCCACACAGTGAAGTCTC...CGGTCGCGCG 50815
      167 LeuProAlaValGluArgValLysProArgIleLysLysSerGlnPr 183
      50814 AGACGAGCAATGAG.....CCCCGAGCAGCGGATGTCGTCGCG 50774
      183 oThrCysLysThrAlaProAspTyrGlyGlnGlyMetAsn..... 196
      50773 CCGTTGCGAAGCTGCGGAAGCCGAGCGAGGCTTCGACCCATGACGCT 50724
      197 ..ProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGlu 212
      50723 GTCTCTGATCCAGTACCGCTCTGCTGCCACGAGATAGTCCGACGCGAA 50674
      213 ProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPh 229
      50673 CCC.....GCCTGCCGCGCGCGGGA..... 50653
      229 eValMetGlnValLysValGlyHisHisThrAlaGluGlyValGlyThra 246
      50652 .....ACAGCCGAGCCAGCTCACC 50631

```

```

      246 snLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeu 262
      50630 ATAGCCGAGCGCCACAGCGTCCCGACGCGCT..... 50599
      263 GlyPheLysValProGlnAlaGlnProAlaLysProAlaLeu.....Ly 277
      50598 .....CCAGACGCTGCGCGCTGCTCTGCGCCCTGCGCCGACG 50561
      277 sSerGluGluThrProValLysLysProGluLysAspGlyArgLysValT 294
      50560 GAGCCACCGCAGCGCCCTTGTCTGCGCCCGCTGTGATCTGTCCA. 50512
      294 hrPhePheGluProSerProGluAspGluAsnGlyThrSerAsnLysAsp 310
      50511 .....GGGCGGACACCAAGATCGGGTGGGCTCATCT..... 50479
      311 GluGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyI 327
      50478 .....CGATGAACAGCGCGGGGCCACCT.....CCAGCAGCGCT 50444
      327 eleuProMetValProGluValAlaGlnAlaValGlyVal.....Serg 342
      50443 TCGCGCGCGCAGCGGAAGCGCACCGCTGCGGAAGTGTGCCGCCAGTA 50394
      342 lngLysHisThrLysAspPheThrArgAlaAlaProAsnProAlaLys 358
      50393 GCTCGACCGAGCTCCGAGCCCGCATCACCCCGCGCTCACCGCTCGAGC 50344
      359 AlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSe 375
      50343 GCATCG.....GCACCGCAGCCGCTCGCGCGCGATCGCT 50309
      375 rProThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyHisValP 392
      50308 CCCAGCGCGCGATCAGCTCTTCGCGCA.....GCGGTCGACCTGCGG 50265
      392 roHisGlyProArgThrArgProSer 400
      50264 GCTATGGCTGCGGACGTCCACCTTCA 50239

```

OM of: US-09-316-048-11 to: EST:* out_format : pfs
Date: Apr 5, 2001 9:12 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09316048/runat_01042001_170038_7731/app_query.fasta_1.1293
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdl
-LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09316048_@CGN1_1_4633 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPRX -WAIT -THREADS=1

Search information block:

Query: US-09-316-048-11

Query length: 487

Database: EST:*

Database sequences: 7991742

Database length: -791223438

Search time (sec): 2268.490000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est76:BE867390	+ 1132.00	1975.59	5.0e-101	1018	BE867390 601442470F1 NIH_MGC_6
gb_est43:AW320695	+ 1006.00	1758.70	6.0e-89	633	AW320695 uo22b10.y1 NCI_CGAP_Ma
gb_est77:BE894295	+ 981.50	1713.50	2.0e-86	752	BE894295 601437666F1 NIH_MGC_72
gb_est74:BE674157	+ 959.00	1676.08	2.4e-84	616	BE674157 7d76a04.x1 NCI_CGAP_Lu
gb_est73:BE613972	+ 915.00	1597.22	5.9e-80	686	BE613972 601504093F1 NIH_MGC_71
gb_est40:AW141293	+ 854.00	1493.11	3.7e-74	504	AW141293 EST291334 Normalized r
gb_est40:AW162002	+ 836.00	1459.47	2.8e-72	594	AW162002 au72c03.x1 Schneider f
gb_est66:BE082712	+ 835.00	1457.31	3.7e-72	615	BE082712 RC2-BT0642-150200-012-
gb_est75:BE781223	+ 826.00	1440.40	3.2e-71	673	BE781223 601469159F1 NIH_MGC_67
gb_est77:BE872137	+ 823.00	1432.42	8.9e-71	852	BE872137 601446238F1 NIH_MGC_65
gb_est76:BE810330	+ 808.50	1410.67	1.5e-69	609	BE810330 MR4-PT0052-220500-206-f
gb_est40:AW163206	+ 808.00	1410.66	1.5e-69	564	AW163206 au97f05.x1 Schneider f
gb_est73:BE617223	+ 790.00	1375.66	1.3e-67	749	BE617223 601441906F1 NIH_MGC_65
gb_est72:BE541462	+ 781.00	1361.00	8.5e-67	673	BE541462 601067913F1 NIH_MGC_10
gb_est76:BE868328	+ 771.50	1341.36	1.1e-65	866	BE868328 601443887F1 NIH_MGC_65
gb_est66:BE089861	+ 770.00	1344.04	7.5e-66	543	BE089861 RC5-BT0708-150300-021-
gb_est4:AA240112	+ 765.50	1337.49	1.7e-65	481	AA240112 my22a02.r1 Barstead mc
gb_est46:AW580332	+ 765.00	1335.78	2.2e-65	517	AW580332 IL2-HT0449-100100-032-
gb_est69:BE300155	+ 764.00	1328.63	5.4e-65	829	BE300155 600944602F1 NIH_MGC_17
gb_est43:AW351909	+ 752.00	1313.36	3.8e-64	494	AW351909 RC0-CT0202-200999-011-
gb_est50:AW877286	+ 732.00	1277.40	3.9e-62	524	AW877286 MR4-PT0052-150200-101-
gb_est4:AA270608	+ 731.50	1277.64	3.7e-62	475	AA270608 va66b02.r1 Soares mous
gb_est40:AW153775	+ 729.50	1270.44	9.4e-62	655	AW153775 f125e06.y1 Sugano Kaw
gb_est44:AWA03841	+ 724.00	1264.63	2.0e-61	466	AWA03841 UI-HF-BK0-abl-g-10-0-0
gb_est3:AA206573	+ 702.00	1223.45	3.9e-59	573	AA206573 zq51f08.r1 Stratagene
gb_est28:AL045575	+ 679.50	1181.63	8.3e-57	690	AL045575 DKFZp434M205_r1 434 (S
gb_est86:U69197	+ 669.00	1152.15	3.7e-55	1801	U69197 U69197 Soares infant br
gb_est51:AW952360	+ 658.00	1143.14	1.2e-54	724	AW952360 EST364430 MAGE reseque
gb_est27:AT958903	+ 636.00	1107.34	1.1e-52	556	AT958903 fd05f07.y1 Zebrafish W
gb_est46:AW581619	+ 632.00	1101.52	2.1e-52	499	AW581619 CM3-PT0033-030100-258-
gb_est85:T31173	+ 631.00	1102.57	2.1e-52	390	T31173 EST28292 Human Brain Hom
gb_est66:BE093820	+ 628.00	1096.48	4.6e-52	418	BE093820 PM0-BT0757-120400-001-
gb_est86:W28253	+ 608.00	1052.53	1.3e-49	893	W28253 44b6 Human retina cDNA r
gb_est49:AW836710	+ 607.00	1058.61	5.9e-50	449	AW836710 OV1-LT0035-070300-101-
gb_est69:BE306047	+ 607.00	1053.14	1.2e-49	725	BE306047 601101723F1 NCI_CGAP_L
gb_est84:R62169	+ 600.00	1046.65	6.9e-48	434	R62169 y120g06.s1 Soares placen
gb_est38:AV705048	+ 589.00	1021.56	6.9e-48	714	AV705048 AV705048 ADB Homo sapi
gb_est69:BE299828	+ 586.50	1015.42	1.5e-47	831	BE299828 600944602F1 NIH_MGC_17
gb_est85:T31172	+ 569.00	993.87	2.4e-46	367	T31172 EST28289 Human Brain Hom
gb_gss27:CNS04FLC	- 560.50	969.28	5.6e-45	850	AL288489 Tetraodon nigroviridis
gb_est3:AA218774	+ 544.00	950.20	6.5e-44	353	AA218774 zq97a02.r1 Stratagene

gb_est75:BE784014	+	536.50	926.02	1.4e-42	921	BE784014 601472491F1 NIH_MGC
gb_est69:BE278193	+	534.00	925.33	1.6e-42	665	BE278193 601180017F1 NIH_MGC
gb_est81:H04110	+	515.50	897.37	5.7e-41	441	H04110 yj48d08.r1 Soares pla
gb_est79:BF016149	+	503.00	873.39	1.2e-39	522	BF016149 uy31e08.y1 NCI_CGAP

seq_name: gb_est76:BE867390

seq_documentation_block:
LOCUS BE867390 1018 bp mRNA EST 27-SEP-2000
DEFINITION 601442470F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846689 5',
mRNA sequence.
ACCESSION BE867390
VERSION BE867390.1 GI:10316166
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE 1 (bases 1 to 1018)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLCM546 row: 1 column: 18
High quality sequence stop: 654.

FEATURES
source Location/Qualifiers
1..1018
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846689"
/tissue_type="NIH_MGC_65"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; site_1: NotI;
site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 295 a 267 c 262 g 194 t
ORIGIN

alignment_scores:
Quality: 1132.00 Length: 286
Ratio: 4.288 Gaps: 5
Percent Similarity: 92.308 Percent Identity: 82.168

alignment_block:

US-09-316-048-11 x BE867390 ..

Align seg 1/1 to: BE867390 from: 1 to: 1018

202 ALAGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrMetLeuLe 218
|||||
3 GCCCAGATCCAGCAGGCAAAAAGAGAGAGAGCCAGCTACACGCTCCT 52
218 uThrgluArgGlyLeuProArgArgArgGluPheValMetGlnValLysV 235
|||||
53 CACAGACCGAGGCTCCCGCCGACAGGAGTGTGATGCAGGTGAAGG 102
235 aAGlyHisHisThrAlaGluGlyValGlyThrAsnLysLysValAlaLys 251
|||||
103 TTGGAACACCACTGCAGAGAGACGCGCACCAACAGAGAGGTGCCAAG 152
252 ArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysValProG1 268

|||||
153 CGCAATGCAGCCGAGACATGCTGGAGATCCTTGTTTCAAGTCCCGCA 202
268 nAlaGlnProAlaLysProAlaLeuLysSerGlnGluLysThrProVal 285
|||||
203 GGGCGAGCCGACCAACCCGCACTCAAGTCAGAGAGAACACACCATAA 252
285 yLysProGlnAspGlyArgLysValThrPhePheGluProSerProGly 301
|||||
253 AGAAACAGGCGATGGAAGAAAGTAACCTTTTGAACCTGGCTGGG 302
302 AspGluAsnGlyThrSerAsnLysAspGluPheArgMetProTyrLe 318
|||||
303 GATGAAATGGAGCTAGTAATAAAGAGATGAGTTCAGGATGCCCTATCT 352
318 uSerHisGlnGlnLeuProAlaGlyIleLeuPrometValProGluValA 335
|||||
353 AAGTCATCAGCAGCTGCTGCTGGAATTCCTCCCATGGTGCCCGAGGTGG 402
335 laGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArg 351
|||||
403 CCCAGGCTGAGAGTTAGTCAAGAGACATCACACCAAGATTTTACCAAG 452
352 AlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgG 368
|||||
453 GCAGCTCCGAATCTGCCAAGGCCACGCTACTGCATGATAGCCCGAGA 502
368 uLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSera 385
|||||
503 GTTGTGCTATGGGGGCACTCGCCACAGCCGAGACCATTTTAAAGATA 552
385 snIleSerSerGlyHisValProHisGlyProArgThrArgProSerGlu 401
|||||
553 ACATCTCTTCAGGCGCCAGCTACCCCATGACCTCTCACGAGACCCCTGAG 602
402 GlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGluTyrLysAs 418
|||||
603 CAACTGGACTATCTTTCCAGAGTCCAGGGATTCAGGTTGAATACAAAGA 652
418 pPheProLysAsnAsnLysAsnGlyCys...ValSerLeuIleAsnCys 434
|||||
653 CTTCCCAAAATCAACCAAGACGAATTCGCTATCTCTTATCAATGGCT 702
434 eSerGlnProProLeuValSerHisGlyIle.GlyLysAspValGluSe 450
|||||
703 CCTCTCAGCCACT CTGATCCACATGTGTTCGCGAGGTTGTGGGAATC 751
450 rCysHis...AspMetAlaAlaLeuAsnIleLeuLysLeu.LeuSerGlu 465
|||||
752 CTGGCAATGATCATGGGTGGGCTGGACATCTTTAAGATGGCTGGCGAA 801
466 LeuAspGlnGlnSerThr.GluMetProArgThrGlyAsnGlyProValS 482
:::|
802 GTGGACCCACAAAGTCAAAAATTTCCCAAGAAACGGGAGGAGCCAGCT 851
482 er 482
||
852 CG 853

seq_name: gb_est43:AW320695

seq_documentation_block:

LOCUS AW320695 633 bp mRNA EST 25-JAN-2000
DEFINITION uc02b10.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:2609851 5'
similar to TR:O95793 O95793 STAUFEN PROTEIN. ;, mRNA sequence.

ACCESSION AW320695
VERSION AW320695.1 GI:6750239

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 633)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

MG1:1023711
Seq primer: -40RP from Gibco
High quality sequence stop: 356.

FEATURES
Location/Qualifiers

1..633
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:2609851"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 176 a 167 c 163 g 127 t
ORIGIN

alignment_scores: Quality: 1006.00 Length: 206
Ratio: 5.005 Gaps: 0
Percent Similarity: 97.573 Percent Identity: 93.204

alignment_block:
US-09-316-048-11 x AW320695 ..
Align seg 1/1 to: AW320695 from: 1 to: 633

268 GlnAlaGlnProAlaLysProAlaLeuLysSerGlnGluLysThrProVa 284
|||||
16 CAGGCGAGCGCTGCCAAGCCAGCACTCAAAATCAGAAAGAGAGAGCTCCAGT 65
284 lLysLysProGlnAspGlyArgLysValThrPhePheGluProSerProG 301
|||||
66 AAAGAAACAGAGAGAGCGAAGAAAGTAACCTTTTGAACCTAGCCCTG 115
301 lyAspGluAsnGlyThrSerAsnLysAspGluPheArgMetProTyr 317
|||||
116 GGGATGAAATGGAAGTACTAGTAACAAGAGCAGAGTTCAGGATGCCCTAT 165
318 LeuSerHisGlnGlnLeuProAlaGlyIleLeuPrometValProGluVa 334
|||||
166 CTTAGCCATCAGCAGCTGCCAGCTGGAATTCCTCCCATGTGTGCCGGAAGT 215
334 lAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThra 351
|||||
216 TGCCGAGGCTGTGGGGTTAGTCAAGGACACACACCAAGATTTCACCA 265
351 rgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaarg 367
|||||
266 GGGCAGCTCCAATCTCGCCCAAGCAACGCTAAGTGCATGATAGCCCGA 315
368 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSe 384
|||||

316 GAGTTGTGTACGGGGGACCTCGCCACAGCCGAGACCATTTTAAGAG 365
384 rAsnIleSerSerGlyHisValProHisGlyProArgThrArgProSerG 401
366 TAACATCTCTTCAGGCCACGTAACCCGATGACCTCGCACTAGACCCCTCTG 415
401 LuGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGluTyrLys 417
416 AGCAACTGTACTACCTTTCAGAGCCAGGATTCAGGTTGAATACAAA 465
418 AspPheProLysAsnAsnLysAsnGluCysValSerLeuIleAsnCysSe 434
466 GATGTTGCCAAGACGACGAGACGAGTGTGTATCTCTCATCAACTGCTC 515
434 rSerGlnProProLeuValSerHisGlyIleGlyLysAspValGluSerC 451
516 CTCACAGCCGGCTCTCGTCAGTCATGGCATCGGCAAGATGTGAAGTCC 565
451 ySHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLeuAsp 467
566 GTCATGATATGGCTGCACCTGACATTTGTAAGCTGCTGTGTGAGTTGAC 615
468 GlnGlnSerThrGluMet 473
616 TCACAGAACACAGATATT 633

seq_name: gb_est77:BE894295

seq_documentation_block:
LOCUS BE894295 752 bp mRNA EST 29-SEP-2000
DEFINITION 601437666F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922529 5',
mRNA sequence.

ACCESSION BE894295
VERSION BE894295.1 GI:10356518
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM744 row: d column: 18
High quality sequence stop: 654.

FEATURES
source

1..752
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922529"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: PCMV-SPORT6; site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 223 a 194 c 192 g 143 t
ORIGIN

alignment_scores: 981.50 Length: 237
Quality: 981.50

Ratio: 4.461 Gaps: 3
Percent Similarity: 92.827 Percent Identity: 84.810
alignment_block:
US-09-316-048-11 x BE894295 ..

Align seg 1/1 to: BE894295 from: 1. to: 752

191 TyrGlyGlnGlyMetAsnProIleSerArgLeuAlaGlnIleGlnGlnAl 207
3 TATGCCCGGCGGATCAATCCGATTAGCCGACTGGCCGACATCCAGAGGC 52
207 aLysLysGluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuP 224
53 AAAAAAGGAGAGAGAGCCAGAGTACAGCTCCTCACAGAGCGAGCCCTCC 102
224 roArgArgArgGluPheValMetGlnValLysValGlyHisHisThrAla 240
103 CGCGCCGCGAGGAGTTGTGTGATGCAGGTGAAGTTGGAACACACACTGCA 152
241 GluGlyValGlyThrAsnLysLysValAlaLysArgAsnAlaGluAs 257
153 GAAGGAGCGGCCACCAACAAGAGGTGGCCAAAGCGCAATGCAGCCGAGAA 202
257 nMetLeuGluIleLeuGlyPheLysValProGlnAlaGlnProAlaLysP 274
203 CATGCTGGAGATCCTTGGTTTCAAAAGTCCCGCAGCGCGACCCACCAAC 252
274 roAlaLeuLysSerGluGluLysThrProValLysLysProGlyAspGly 290
253 CCGCAGCTCAGTCAAGAGAGAGAGACACCCATTAAGAAGAACACAGG. GATGGA 301
291 ArgLysValThrPhePheGluProSerProGlyAspGluAsnGlyThrSe 307
302 AGAAAAGTAACTTTTGTGAACCTGGCTCTGGGGATGAATAATGGGACTAG 351
307 rAsnLysAspGluGluPheArgMetProTyrLeuSerHisGlnGlnLeuP 324
352 TAATAAAGAGATGAGTTCAAGATGCCCTTATCTAAGTCATCAGACAGTGC 401
324 roAlaGlyIleLeuProMetValProGluValAlaGlnAlaValGlyVal 340
402 CTGCTGGAATTCTTCCCATGTGTCGCCGAGGTGCGCCAGGCTGTAGAGATT 451
341 SerGlnGlyHisHisThrLysAspPheThrArgAlaAlaProAsnProAl 357
452 AGTCAAGGACATCACACCAAGAATTTTACAGAGACAGCTCCGAATCCTGC 501
357 aLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyT 374
502 CAAGGCCACGGTAACTGCCATGATAGCCCGAGAGTTGTAGTATGGGGCA 551
374 hrSerProThrAlaGluThrIleLeuLys.SerAsnIleSerSerGlyHis 390
552 CCTCGCCCGACAGCCGAGATCATTTTAAACGAATAACATCTCTCAGGCCA 601
390 sValProHisGlyProArgThrArgPro...SerGluGlnLeuTyrTyrL 406
602 CGTACCCCATGCACTCTCACGAGATCCTCTGAGCAATGAGACTATCTTTC. 650
406 euSerArgAlaGlnGlyPheGlnValGluTyrLysAspPheProLysAsn 422
651 CAGAGTCCAGGATTCGGGGTTGATCAAAAGACTTTCCCAAAAAA 695
423 AsnLysAsn 425
696 CAGAGGAAT 704

seq_name: gb_est74:BE674157

seq_documentation_block: 616 bp mRNA EST 08-SEP-2000
LOCUS BE674157
DEFINITION 7d76a04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3278862 3'

```

similar to TR:O95793 O95793 STAUFEIN PROTEIN. ;, mRNA sequence.
ACCESSION BE674157
VERSION BE674157
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 616)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 469.
FEATURES
Source
1. 616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3278862"
/clone_lib="NCI-CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 179 a 153 c 139 g 141 t 4 others
ORIGIN
alignment_scores:
Quality: 959.00 Length: 202
Ratio: 4.969 Gaps: 0
Percent Similarity: 95.545 Percent Identity: 90.099
alignment_block:
US-09-316-048-11 x BE674157 ..
Align seg 1/1 to: BE674157 from: 1 to: 616
281 LysThrProValLysLysProGlyAspGlyArgLysValThrPhePheG1 297
|||||:::|||||
8 AAGACACCCATAAAGAAACCAAGGGGATGGAAGAAAGTAACCTTTTGA 57
297 uProSerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheA 314
||||:::|||||
58 ACCTGGCTCTGGGGATGAATAATGGGACTAGTAATTAAGAGAGATGAGTTCA 107
314 rGmetProTyrLeuSerHisGlnGlnLeuProAlaGlyTleLeuPromet 330
|||||
108 GGATGCCCTATCTAATGATCATCAGCAGCTGCCCTGCTGCAATTCCTCCATG 157
331 ValProGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisThrLy 347
|||||
158 GTGCCCGAGGTGGCCAGGCTGTAGAGATTAGTCAAGACATCACACCAA 207

```

```

347 sasprhethrargalaalaproasnproalalysalathrvalthralam 364
|||||
208 agatTTTACCAGGGCAGCTCCGAATCTGCCAAGGCCACGGTAAC TGCCA 257
|||||
364 etllealAargGluleuLeuTyrglyGlyThrSerProthralagluThr 380
|||||
258 TGATAGCCCCGAGAGTTGTTGATGGGGCCACCTCGCCACAGCGAGACC 307
|||||
381 lleuLysSerAsnIleSerSerGlyHisValProHisGlyProArgTh 397
|||||
308 ATTAAAGAATAACATCTCTTCAGGCCACGATACCCCATGACCTCTCAC 357
|||||
397 rArgProSerGluGlnLeuTyrrLeuSerArgAlaGlnGlyPheGlnV 414
|||||
358 GAGACCCCTGTGAGCAACTGACATCTTTCAGAGTCCAGGATTCAGG 407
|||||
414 aIGluTyrrLysAspPheProLysAsnLysAsnGluCysValSerLeu 430
|||||
408 TTGAATACAAAGACTTCCCCAAAACAACAAGAAGCAATTTGTATCTCTT 457
|||||
431 IleAsnCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAs 447
|||||
458 ATCAATTGCTCTCTCANCACCTCTGTGATCAGCCATGGTATCAGCAAGGA 507
|||||
447 pValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuS 464
|||||
508 TGTGGAGTCTGCGCATGATATGGCTGCGCTGAACATCTTAGAGTTGCTGT 557
|||||
464 erGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyPro 480
|||||
558 CTGAATGTGACCAACAANTACAGAGATGCCAAGAACANGANACAGACCA 607
|||||
481 ValSer 482
:::||||
608 ATGTCT 613

seq_name: gb_est73:BE613972

seq_documentation_block:
LOCUS BE613972 686 bp mRNA EST 24-AUG-2000
DEFINITION 601504093F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5',
mRNA sequence.
ACCESSION BE613972
VERSION BE613972.1 GI:9895569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 686)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LCM700 row: j column: 24
High quality sequence stop: 675.
Location/Qualifiers
1. 686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3905783"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"

```


/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

BASE COUNT 226 a 156 c 196 g 108 t
ORIGIN

alignment_scores: Length: 230
 Quality: 915.00 Gaps: 4
 Ratio: 4.378
Percent Similarity: 90.870 Percent Identity: 84.348

alignment_block:
US-09-316-048-11 x BE613972 ..

Align seg 1/1 to: BE613972 from: 1 to: 686

```
62 HisAspAlaProAlaArgAlaLeuArgThrLeuGlnSerGluProLeuPr 78
|||||  |||:|||||  |||:|||||
3 CAGGATGCTGCTGCCAAGCGTTGAGGATCTTCGAGAATGAGCCCTGCC 52
78 oGluArgLeuGluValAsnGlyArgGluAlaGluGluGluAsnLeuAsnL 95
|||||  |||:|||||  |||:|||||
53 AGAGAGCGCTGGAGGTTGAATGAGAGCAATCCGAGAAGAAATCTCAATA 102
95 ySSerGluLeSerGlnValPheGluIleAlaLeuLysArgAsnLeuPro 111
|||||  |||:|||||  |||:|||||
103 AATCTGAATTAAGTCAAGTGTGAGATTGCACCTTAACGGAACCTTGCC 152
112 ValAsnPheGluValAlaArgGluSerGlyProProHisMetLysAsnPh 128
|||||  |||:|||||  |||:|||||
153 GTGAATTTCGAGGTGGCCCGGAGAGTGGCCACCCACACATGAAGAACTT 202
128 eValThrArgValSerValGlyGluPheValGlyGluGlyLys 145
|||  |||  |||:|||||  |||:|||||
203 GTGACCAAGGTTTCGGTGGGA...GTTGTGGGGAAGGTGAAGGAAAA 249
145 erLysLysIleSerLysLysAsnAlaAlaArgAlaValLeuGluGlnLeu 161
|||||  |||:|||||  |||:|||||
250 GCAAGAGATTTCAAAGAAAAATGCCCATAGCTGTCTTGAGAGAGCTG 299
162 ArgArgLeuProProLeuProAlaValAluArgValLysProArgIleLy 178
:::  |||:|||||  |||:|||||
300 AAGAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAGCCTTAGAATCAA 349
178 sLysLysSerGlnProThrCysLys....ThrAlaProAspTyrGlyG 193
|||||  |||:|||||  |||:|||||
350 AAAGAAAACAAAACCATAGTCAAGCCACAGACAAGCCAGAAATATGGCC 399
193 lnglYmetAsnProIleSerArgLeuAlaGlnIleGlnAlaLysLys 209
|||||  |||:|||||  |||:|||||
400 AGGATC.AATCCGATTAGCCGAGCTGCCAGATCCAGACAGCAAAAAAG 448
210 GluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgAr 226
|||||  |||:|||||  |||:|||||
449 GAGAAGAGAGCCAGAGTACAGCTCCTCACAAGAGAGCGCTCCCGCCCG 498
226 gArgGluPheValMetGlnValLysValGlyHisHisThrAlaGluGly 243
|||||  |||:|||||  |||:|||||
499 CAGGAGTTGTGATGCAGGTGAAGTTGGAACAACACACTGCAGAAAGAA 548
243 aIglYThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeu 259
:::  |||:|||||  |||:|||||
549 CGGGCACCAACAAGAGGTGGCCAGCGCAATGCAGCCGAGAACATGCTG 598
260 GluIleLeuGly.PheLysValProGlnAlaGlnProAlaLysProAlaL 276
|||||  |||:|||||  |||:|||||
599 GAGATCCTGTGTTTCAAAAGTCCCGAGGCGCAGCCACCAAAAC...CG 645
276 euLysSerGluGluLysThrProValLysLysProGly 288
|||||  |||:|||||  |||:|||||
646 TCAAGTCAGAGAGAGACACACCATTAAGAAACAGAGGA 683
```

seq_name: gb_est40:AW141293

seq_documentation_block:

LOCUS AW141293 504 bp mRNA EST 30-OCT-1999
DEFINITION EST291334 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RGIBK57 5' end similar to maternal effect protein STRUFEN, mRNA
sequence.

ACCESSION AW141293 GI:6161078
VERSION AW141293.1
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 504)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
ATCC

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@ligr.org
For clone availability, additional sequence and expression
information related to this EST please check the TIGR Rat Gene
Index (<http://www.tigr.org/tdb/rgi/html>). To order a clone
contact the ATCC (<http://www.atcc.org/atcc.html>).
Seq primer: M13 Reverse.

FEATURES
source Location/Qualifiers

1..504
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="RGIBK57"
/clone_lib="Normalized rat brain, Bento Soares"
/note="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 138 a 143 c 117 g 106 t
ORIGIN

alignment_scores: Length: 167
 Quality: 854.00 Gaps: 0
 Ratio: 5.145
Percent Similarity: 99.401 Percent Identity: 97.006

alignment_block:
US-09-316-048-11 x AW141293 ..

Align seg 1/1 to: AW141293 from: 1 to: 504

```
303 GluAsnGlyThrSerAsnLysAspGluGluPheArgMetProTyrLeuSe 319
|||||  |||:|||||  |||:|||||
3 GAAATGGAACCCCTAATTAAGAGAGATGAGTTCAAGATGCCCTATCTTAG 52
319 rHisGlnGlnLeuProAlaGlyIleLeuPrometValProGluValAlaAg 336
|||||  |||:|||||  |||:|||||
53 CCATCAGCAGCTGCCAGCCGGGATTCTCCCATGCTGCCAGAGGTTGCC 102
336 lnaLavaGlyValSerGlnGlyHisHisThrLysAspPheThrArgAla 352
|||||  |||:|||||  |||:|||||
103 AGGCTGTGGGGTTAGTCAAGAGACACCATACCAAGATTTCGCCAGGGCA 152
353 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 369
|||||  |||:|||||  |||:|||||
153 GCTCCAAATCCTGCCAAGGCCACGCGTAACCTGCCATGATAGCCCGAAATT 202
369 uLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSerAsnI 386
|||||  |||:|||||  |||:|||||
```

203 GTTGTACGGGGACCTCGCCCCACAGCCGAGACCATTTTAAGAGTAACA 252
386 l e s e r S e r G l y H i s V a l P r o H i s G l y P r o A r g T h r a r g P r o S e r G l u G l n 402
253 T C T C T T C A G G A C A T G T A C C C A T G A C C T C G C A C A G A C C C T C T G A G C A A 302
403 L e u T y r T y r L e u S e r A r g A l a G l n G l y P h e G l n V a l G l u T y r L y s A s p H 419
303 C T G T A C T A C C T T T C C A G A C C C A G G A T T C C A G G T T G A A T A C A A G A C T T 352
419 e P r o L y s A s n A s n L y s A s n G l u C y s V a l S e r L e u I l e A s n C y s S e r S e r G 436
353 C C C C A G A G A C A C A C A G A G A G T G T A T C T C T C A T C A A C T G C T C T T C A C 402
436 I n P r o P r o L e u V a l S e r H i s G l y I l e G l y L y s A s p V a l G l u S e r C y s H i s 452
403 A G C C A C C T C T C G T C A G C C A T G C G C A G A G A T G T G A G T C C T G T C A T 452
453 A s p M e t A l a A l a L e u A s n I l e L e u L y s L e u L e u S e r G l u L e u A s p G l n G l 469
453 G A T A T G G C T G C A C T G A A C A T C T T A A G C T G C T G T C T G A G T T G G A C C A A C A 502
469 n 469
503 G 503
seq_name: gb_est40:AW162002
seq_documentation_block:
LOCUS AW162002 594 bp mRNA EST 09-NOV-1999
DEFINITION au72c03.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2781796 3' similar to TR:O95793 O95793 STAUFEN PROTEIN. ;
mRNA sequence.
ACCESSION AW162002
VERSION AW162002.1 GI:6301035
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 594)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: au72c03.y1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 464.
FEATURES
source
1. 594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2781796"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:

5'-GAGAGAGAGAGAGAGCTCAAGATCCTTAATTAATTAATCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGACTCGAGTCTTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
science Park, Trieste, Italy)."
BASE COUNT 130 a 152 c 123 g 188 t 1 others
ORIGIN
alignment_scores:
Quality: 836.00 Length: 178
Ratio: 4.889 Gaps: 0
Percent Similarity: 96.067 Percent Identity: 89.888
alignment_block:
US-09-316-048-11 x AW162002/rev ..
Align seg 1/1 to reverse of: AW162002 from: 1 to: 594
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
534 ATGTATAAGCCTGTGTGACCTTACTCTCGATGCAGTCCACCTATACTA 485
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
484 CAACATGAGAGGAGGTGCTTATCCCGGAGGACTTTTACCATTTCAG 435
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
|||||
434 TTCACACCTTACTTTATCAAGTGAACCTTCTGTGGGAGACAGCAATT 385
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
384 AATGCCAAAGAAAGACAGACAGAGCTGCGAAACAGATGCTGTCGCAA 335
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
:|||||
334 AGCGTTGAGGATCTTGCACAATGAGCCCCCTGCCAGAGAGGTGAGGTGA 285
84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluLysSerGln 100
|||||
284 ATGGAAGAGAAATCCGAGAAGAAATCTCAATTAATCTGAATAATAGTCAA 235
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
234 GTCTTTGAGATTGCACCTTAACGGAACCTTGCTGTGAATTTGAGAGTGGC 185
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
|||||
184 CCGGAGAGAGTGGCCCAACCCACATGAAGAACTTTGTGACCAAGGTTTCG 135
134 aGlyGluPheValGlyGluGlyGluGlyLysSerLysLysIleSerLys 150
|||||
134 TTGGGGAGTTGTGTGGGGGAAGGTGAAGGGAAAGCAAGAGATTTCAAAG 85
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProLe 167
|||||
84 AAAAATGCCCGCATAGCTGTCTTGAGGAGCTGAAGAAGTTACCGCCCT 35
167 uProAlaValGluArgValLysProArgIleLys 178
|||||
34 GCCTGCAGTTGAACGAGTAAGAAGCTAGAAAAAAA 1
seq_name: gb_est66:BE082712
seq_documentation_block:
LOCUS BE082712 615 bp mRNA EST 12-JUN-2000
DEFINITION RC2-BT0642-150200-012-a08 BT0642 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE082712

VERSION BE082712.1 GI:8473017
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 615)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612=RC2-BT0642-150200-012-a086t3=2000-02-15&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 75
High quality sequence stop: 609.
FEATURES
source
1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0642"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 110 a 159 c 153 g 193 t
ORIGIN
alignment_scores:
Quality: 835.00 Length: 187
Ratio: 4.639 Gaps: 1
Percent Similarity: 96.257 Percent Identity: 86.096
alignment_block:
US-09-316-048-11 x BE082712/rev ..
Align seg 1/1 to reverse of: BE082712 from: 1 to: 615
144 lysSerLysLysIleSerLysLysAsnAlaAlaAArgAlaValLeuGluc1 160
|||||
613 AAAAGCAAGAGATTCAAGAAAAATGCCGCATAGCTGTTCTTGAGGA 564
160 nleuAArgLeuProProleuProAlaValAluArgValysProArg1 177
:|||||
563 GCTGAAGAAGTTACCGCCCTCGCTGAGTAGACGAGTAAAGCCTAGAA 514
177 leLysLysLysSerGlnProThrCysLys.....ThrAlaProAspTyr 191
|||||
513 TCATAAAGAAACCAACCATAGTCAAGCCACAGACAAGCCAGAAATAT 464
192 GlyGlnGlyMetAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLy 208
|||||

463 GGCCAGGGGATCAATCCGATTAGCCGACTGGCCAGATCCAGCAGCAAA 414
208 sLysGluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuProA 225
|||||
413 AAAGGAGAAGAGAGCCAGAGTACACGCTCCTCACAGAGCGAGCCTCCGC 364
225 rGArgArgGluPheValMetGlnValLysValGlyHisHisThrAlaGlu 241
|||||
363 GCCGAGGAGAGTTGTGTGATGACAGGTGAAGGTTGGAACACACACTGCAGAA 314
242 GlyValGlyThrAsnLysLysValAlaLalysArgAsnAlaAlaGluAsnMe 258
|||||
313 GGAACGGGCACCAACAAGAAGGTGGCCAAGCCCAATGCACGCCGAGAACAT 264
258 lleuGluIleLeuGlyPheLysValProGlnAlaGlnProAlaLysProA 275
|||||
263 GCTGAGATCCTTGGTTCAAGTCCCGCAGCGCCAGCCACCAACCCG 214
275 laLeuLysSerGluGluLysThrProValLysLysProGlyAspGlyArg 291
|||||
213 CACTCAAGTCAGAGAGAGAACACCCATAAAGAAACACAGGGGATGAAGA 164
292 LysValThrPheGluProSerProGlyAspGluAsnGlyThrSerAs 308
|||||
163 AAAGTAACCTTTTGAACCTGGCTCTGGGATGAATAATGGAGACTAGTAA 114
308 nLysAspGluGluPheArgMetProTyrLeuSerHisGlnGlnLeuProA 325
|||||
113 TAAAGAGATGAGTTCAGAGATGCCCTATCTAAGTCATCACCATCTTCCTG 64
325 laGlyIleLeu 328
|||||
63 CTGGACTTCTT 53
seq_name: gb_est75:BE781223
seq_documentation_block:
LOCUS BE781223 673 bp mRNA EST 20-SEP-2000
DEFINITION 601469159F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3872247 5', mRNA sequence.
ACCESSION BE781223
VERSION BE781223.1 GI:10202421
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM613 row: e column: 16
High quality sequence stop: 644.
FEATURES
source
1..673
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3872247"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT."

Average insert size 1.75 kb. Library constructed by Life Technologies.
BASE COUNT 241 a 156 c 178 g 98 t
ORIGIN

alignment_scores: Quality: 826.00 Length: 227
Ratio: 4.325 Gaps: 5
Percent Similarity: 84.141 Percent Identity: 76.652

alignment_block:
US-09-316-048-11 x BE8781223 ..

Align seg 1/1 to: BE8781223 from: 1 to: 673

94 AsnlySerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLe 110
1 AATAAATCTGAATAAGTCAAGGTTTGAGATTGCACCTTAACGGAACCTT 50
110 uProValAsnPheGluValAlaArgGluSerGlyProProHisMetLysA 127
51 GCCTGTGAATTTCGAGGTGGCCCGGAGAGTGGCCACCCACACATGAAGA 100
127 snPheValThrArgValSerValGlyGluPheValGlyGluGlyGluGly 143
101 ACTTTGTGACCAGGTTTCGGTTGGGAGTTTGTGGGGAAGGTGAAGGG 150
144 LysSerLysLysIleSerLysLysAsnAlaAlaArgAlaValLeuGluG1 160
151 AAAAGCAGAAGATTTCAAAGAAAAATGCCGCCATAGCTGTTCTTGAGGA 200
160 nLeuArgArgLeuProProLeuProAlaValGluArgValLysProArgI 177
201 GCTGAAGAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAGCCTTAGAA 250
177 leLysLysLysSerGlnProThrCysLys....ThrAlaProAspTyr 191
251 TCAAAAAAGAAAAACAAACCATAGTCAAGCCACAGACAAAGCCCAAGATAT 300
192 GlyGlnGlyMetAsnProIleSerArgLeuAlaGlnIleGlnAlaLys 208
301 GGCCAGGGAGTCAATCCGATTAGCCGACTGGCCACAGATCCAGCAGCAAA 350
208 sLysGluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuProA 225
351 AAAGAGAGAAGGACCAAGAGTACACGCTCCTCACAAGAGAGCCCTCCCGC 400
225 rgArgArgGluPheValMetGlnValLysValGlyHisHisThrAlaGlu 241
401 GCCGACGAGGAGTTGTGATGCAGGTGAAGGTTGGAACACACACTGCAGAA 450
242 GlyValGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMe 258
451 GGAACGGGACCAACAAGAAGGTGGCCCAAGCCAATGCAGCCGAGAACAT 500
258 tLeuGluIleLeuGly.PheLysValProGlnAlaGlnProAlaLys... 273
501 GCTGAGATCCTTGCTTTCAAAGTCCGACGCGCAGCCACCAAAACC 550
274ProAlaLeuLysSerGluGlu 281
551 CAAAGACACCCCGCAAAAAAATAAAGAAAGAAACATATAAATAA 600
281 sThrProVal.....LysLysProGlyAspGlyArgL 292
601 AACACAATAAACAACCAACATTAACAACAATAAAGGGGGGGGGCCGCC 650
292 ysValThrPhePheGluProSerProGly 301
651 AA.....AAAAGTCCCGCGGC 667
seq_name: gb_est77:BE872137

seq_documentation_block:

LOCUS BE872137 852 bp mRNA EST 27-SEP-2000
DEFINITION 601446238F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850496 5',
mRNA sequence.
ACCESSION BE872137 GI:10320913
VERSION BE872137.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM556 row: k column: 09
High quality sequence stop: 637.

FEATURES
source Location/Qualifiers
1..852

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3850496"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 275 a 215 c 218 g 144 t
ORIGIN

alignment_scores: Quality: 823.00 Length: 252
Ratio: 3.846 Gaps: 9
Percent Similarity: 84.921 Percent Identity: 73.810

alignment_block:
US-09-316-048-11 x BE872137 ..

Align seg 1/1 to: BE872137 from: 1 to: 852

205 GlnGlnAlaLysLysGluLysGluProGluTyrMetLeuLeuThrGluAr 221
1 CAGCAGGCAAAAAAGAGAAGAGCCAGAGTACACGCTCCTCAGAGAGCG 50
221 gGlyLeuProArgArgArgGluPheValMetGlnValLysValGlyHis 238
51 AGGCTCCCGCGCGCAGGAGTTGTGATGCAGGTGAAGGTTGGAACC 100
238 iSThrAlaGluGlyValGlyThrAsnLysLysValAlaLysArgAsnAla 254
101 ACACTGCAGAGGAACGGGCACCAACAAGAGGTGGCCAAAGCCAATGCA 150
255 AlagLuAsnMetLeuGluIleLeuGlyPheLysValProGlnAlaGlnPr 271
151 GCCGAGAACATGCTGAGATCCTTGTTCAAAAGTCCCGCAGCGCAGGCC 200
271 oAlaLysProAlaLeuLysSerGluGluLysThrProValLysLysProg 288
201 CACCAAAACCGCACTCAAGTCAAGAGAGAAGACACCCATATAAGAAACAG 250

288 LyspGlyArg...LysValThrPhePheGluProSerProGlyAsp... 302
|||||
251 GGGATGGAAGAACACAGTCAACCTTTTGAACCTGGCTCTGGGCATGC 300
303 GluAsnGlyThrSer.....AsnLysAspGluLuphe.ArgMetPro. 316
:::|||||
301 AACATGGGCACCTCAGTATCACAAGCAGCATGAGTCCAGATGCCCTT 350
317 ..TyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValPro 332
||| |||||
351 CATCTAAGTCCATCCAGACGCTGCTGGAATCTTCCATGGTGCC 400
333 GluValAlaGlnAlaValAlaGlyValSerGlnGlyHisHisThrLysAspPh 349
|||||
401 GAGGTGCGCCAGGCTGTAGAGTTAGTCAAGGACATCACACCAAGATTT 450
349 eThrArg...AlaAlaProAsnProAlaLysAlaThrValThrAlaMet. 364
|||||
451 TACCAGGGACACAGCTCCGACATCCTGACAAGGCCACGGTAACCTGACCTGC 500
365 .IleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrI 381
:::|||||
501 ACTAGCCCGAGAGTCCGTAGTATGGGGCACCCTGCCACACAGCCGAGACCA 550
381 IeLeuLysSerAsnIleSerSerGlyHisValProHisGlyProArgThr 397
|||||
551 TTTTACAGATATACATCTCTTTCAGGCCACGTAACCCATGGAACCTCTCACG 600
398 ArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGln.V 414
|||||
601 AGACCTCTGAGCAACTGCACTATCTTACCAGAGTCCAGGCATACCAAG 650
414 aLgIuTyrLysAspPheProLysAsnAsnLysAsnGluCysValSer.Le 430
||:|||||
651 TTGACTACAAAGAGACTTCCCAAAAGCAAAAGAACGAATAGTATCTCTT 700
430 uIleAsnCysSerSerGlnProLeuValSerHisGlyIleGlyLys 446
|||||
701 AATCAATTGGTACTCTCAGCAACGCTGTATCAGCAGAGAATCGGGAAG 749
seq_name: gb_est76:BE810330

seq_documentation_block: 609 bp mRNA EST 21-SEP-2000
LOCUS BE810330
DEFINITION MR4-PT0052-220500-206-f06 PT0052 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE810330
VERSION BE810330.1 GI:10242524
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=MR4-PT0052-220
500-206-f06&tl3=2000-05-22&tl4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 589.
Location/Qualifiers

FEATURES
source

1.609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="PT0052"
/dev_stage="Adult"
/note="Organ: pnet; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 173 a 153 c 145 g 138 t
ORIGIN

alignment_scores: Quality: 808.50 Length: 185
Ratio: 4.673 Gaps: 2
Percent Similarity: 93.514 Percent Identity: 89.189

alignment_block:

US-09-316-048-11 x BE810330 ..

Align seg 1/1 to: BE810330 from: 1 to: 609

304 AsnGlyThrSerAsnLysAspGluGlu.PheArgMetProTyrLeuSerH 320
:::|||||
8 CATGGGACTAGTAATAAG..AGCATGAGTACAGATGCTTATCTAAGTC 54
320 IsgInGlnLeuProAlaGlyIleLeuProMetValProGluValAlaGln 336
|||||
55 ATCAGCAGCTGCTGCTGA.ATTCTTCCATGCTGCCGAGGTCGCCAG 103
337 AlValGlyValSerGlnGlyHisHisThrLysAspPheThrArgAlaAl 353
|||||
104 GCTGTAGAGTTAGTCAAGGACATCACACCAAGATTTTACAGGGCAGC 153
353 aProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuL 370
|||||
154 TCCGAATCTGCGCAGGGCCACGGTAACTGCCATGATAGCCCGAGAGTAGT 203
370 eUTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSerAsnIle 386
|||||
204 GGTATGGGGCACCTCGCCACAGCCGAGACCATTTTAAGAATAACATC 253
387 SerSerGlyHisValProHisGlyProArgThrArgProSerGluGlnLe 403
|||||
254 TCTTCAGGCCACGATACCCCATGGACCTCTCACGAGAGACCCTTGAGCAACT 303
403 uTyrTyrLeuSerArgAlaGlnGlyPheGlnValGluTyrLysAspPhe 420
| |||||
304 GGACTATCTTTCAGAGTCCAGGGATTCAGGTTGAATACAAAGACTTCC 353
420 rOLysAsnAsnLysAsnGluCysValSerLeuIleAsnCysSerSerGln 436
|||||
354 CCAAAACACAAAGAAGCAATTTGTATCTTATCAATGCTCCTCTCAG 403
437 ProProLeuValSerHisGlyIleGlyLysAspValGluSerCysHisAs 453
|||||
404 CCACCTCTGTATCAGGCATGTGATCGCAAGATGTGTGAGTCTGACCATGA 453
453 pMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLeuAspGlnGlns 470
|||||
454 TATGGCTGCGCTGAACATCTTAAGTTGCTGTCTGAGTAGGACCAACAAA 503

470 erThrGlMetProArgThrGlyAsnGlyProValSerAlaCysGlyArg 486
|||||
504 GTACAGAGATGCCAAGAACAGAAACGACCAATGTCTGTGTGGAGG 553
487 Cys 487
|||
554 TGC 556

seq_name: gb_est40:AW163206

seq_documentation_block:

LOCUS AW163206 564 bp mRNA EST 09-NOV-1999
DEFINITION au97f05.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2784225 3' similar to TR:O95793 O95793 STAUFEN PROTEIN. ;,
mRNA sequence.

ACCESSION AW163206
VERSION AW163206.1 GI:6302239
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 564)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project

TITLE Unpublished (1997)
JOURNAL
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from GIBCO
High quality sequence stop: 455.

FEATURES
Location/Qualifiers
1..564

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2784225"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGACTCAAGATCTTAATTAAATTATATCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGACTCGAGTTTCTTTTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
BASE COUNT 125 a 147 c 117 g 175 t
ORIGIN

alignment_scores:
Quality: 808.00 Length: 178
Ratio: 4.838 Gaps: 0
Percent Similarity: 93.820 Percent Identity: 88.202

alignment_block:

US-09-316-048-11 x AW163206/rev ..
Align seg 1/1 to reverse of: AW163206 from: 1 to: 564

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
534 ATGTATTAGCCTGTGACACCTTTACTCTCGGATGCAGTCCACCTATACTA 485
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|
484 CCACATGAGAGAGAGTGCCTTATCCCCGAGGTACTTTTACCATTTCAG 435
34 alProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50
|||||
434 TTCCACCTTACTTATTCAGAGTGAACCTTCTGTGGAGAGACAGCAATT 385
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
|||||
384 AATGCCAAAGGAAGCAAGACAGAGGCTGCCAAGACGATGCTGCTGCCAA 335
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuValAla 84
:|||||
334 AGCGTTGAGGATCCCTGCAGATGAGCCCTGCCAGAGAGGCTGAGGTGA 285
84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluLeuSerGln 100
|||||
284 ATGGAAGAGAATCCGAAGAAGAAATCTCAATTAATCTGAATAATAGTCAA 235
101 ValPheGluLeuAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
234 GTGTTTGAGATTGCACTTAACGGAACCTTGCCTGTGAATTCGAGGTGCC 185
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSer 134
|||||
184 CCGGAGAGAGTGGCCCAACCCACATGAAGACTTTGTGACCAAGCTTTCGG 135
134 alGlyGluPheValGlyGluGlyGlyLysSerLysLysLysLys 150
|||||
134 TTGGGAGTTTGTGGGGGAAGGTGAAGGGAAGAAAGCAAGAGATTTCAAAG 85
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProPro 167
|||||
84 AAAAATGCCCGCATAGCTGTTCTTGAGGAGCTGAAGAAGTTACCGCCCT 35
167 uProAlaValAlaArgValLysProArgLys 178
|||||
34 GCCTGCAGTTGAACGAGTAAGCCTAGAAAAAAA 1

seq_name: gb_est73:BE617223

seq_documentation_block:
LOCUS BE617223 749 bp mRNA EST 24-AUG-2000
DEFINITION 601441906F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846265 5',
mRNA sequence.

ACCESSION BE617223
VERSION BE617223.1 GI:9888161
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 749)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

Align seg 1/1 to: BE541462 from: 1 to: 673

```
167 LeuProAlaValAlaGluArgValLysProArgIleLys.LysLysSergLup 183
|||||
2 CTGCTGTCAGGTGAACGAGTAAGCCTAGATCAATAAAGAAAACAAAC 51
183 roThrCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnPro 197
|||
52 CCATAGTCAAGCCACAGACAGACCAGAAATATGGCCAGGGATCAATCCG 101
198 IleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProG1 214
|||||
102 ATTAGCCGACTGGCCGATCCAGCAGCAAAAGAGAGAGAGAGAGAGAGAG 151
214 uTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValM 231
|||||
152 GTACACGCTCCTCACAGACGAGGCTCCCGCCGCGCAGGAGTT.GTGA 200
231 etGlnValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLys 247
|||||
201 TGCAGGTGAAGGT.GGAACACACACTGCAGAGGAACGGGCACCAACAAG 249
248 LysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIle.LeuGlyP 264
|||||
250 AAGGTGGCCAGCGCAATGCAGCCGAGACATGCTGGAGATCCTTGGGTT 299
264 helysValProGlnAlaGlnProAlaLys..ProAlaLeuLysSergLup 280
|||||
300 TCAAGTCCCGCAGCGCAGCCACCAAGCCCGCACGTCAGTCAGAGG 349
280 LuLysThrProValLysLysProGlyAspGlyArgLysValThrPhePhe 296
|||||
350 AGAAGACACCCATMAAGAAACAGGGGATGGAAGAAAAGTAACCTTTT 399
297 GluProSerProGlyAspGlu.AsnglyThrSerAsnLysAspGluLup 313
|||||
400 GAACCTGGCTCTGGGATGAACATGGGACTAGTAATAAAGAGATGAGT 449
313 heargMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPro 329
|||||
450 TCAGGATGCCTATCTAAGTCATCAGCAGCTGCCTGCGAATTCCTCCC 499
330 MetValProGluValAlaGlnAlaValGlyValSergGlnGlyHisHis.T 346
|||||
500 ATGGTGGCGAGGTCCGCCAGGCTGTAGAGTTAGTCAAGGACATCACAT 549
346 hrlYsAspPheThrArgAlaAlaProAsnProAlaLys.AlathValTh 362
|||||
550 CCAAGATTTTTACCAAGGGCAGCTCCGAATCTGCCAAGGGCCACGGTAAC 599
362 rAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThr.Ala 378
|||||
600 TGCATGATTAGCCCGAGAGTTGTGTATGGGGCACCTCGTCCAATAGCGG 649
379 GluThrIleLeuLysSer 384
|||||
650 AAGACCATTTTACAGAAAT 667
seq_name: gb_est76:BE868328
```

seq_documentation_block: 866 bp mRNA EST 27-SEP-2000
LOCUS BE868328 601443887F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848252 5',
DEFINITION mRNA sequence.

ACCESSION BE868328
VERSION BE868328.1 GI:10317104
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 866)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L1CM550 row: m column: 21
High quality sequence stop: 605.
Location/Qualifiers

FEATURES
source 1..866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3848252"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 288 a 229 c 224 g 125 t
ORIGIN

alignment_scores: Quality: 771.50 Length: 254
Ratio: 3.605 Gaps: 6
Percent Similarity: 84.252 Percent Identity: 73.228

alignment_block:
US-09-316-048-11 x BE868328 ..

Align seg 1/1 to: BE868328 from: 1 to: 866

```
149 SerLysLysAsnAlaAlaArgAlaValLeuGlnLeuArgArgLeuPr 165
|||||
3 TCAAGAAAAATGCCGCATA.GCTGTTCTTGAGGAGCTGAAGAAGTTACC 51
165 oProLeuProAlaValAlaGluArgValLysProArgIleLysLysSerg 182
|||||
52 GCCCTGCTGCTGAGTTGAACGAGTAAGCCTAGAAATCAAAAAGAAACAA 101
182 InPro..ThrCys..LysThrAlaProAspTyrGlyGlnGlyMetAsn 197
|||||
102 AACCATAGTCAAGCCACAGACAGCCAGAAATATGGCCAGGATC.AATC 150
197 roIleSerArgLeuAlaGlnIleGlnGlnAla.LysLysGluLysGluPr 213
|||||
151 CGATTAGCCGACTGGCCAGATCCAGCAGCAAAAAAGAGAGAGAGGCC 200
213 oGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPhe 230
|||||
201 AGAGTACACGCTCCTCACAGAGCGAGGCTCCCGCGCGCAGGAGTTTG 250
230 aI.MetGlnValLysValGlyHisHisThrAlaGluGlyValGlyThrAs 246
|||||
251 TAGATGACAGGTGAAGTTGGAACACACACTGCAGAGGAAGAGGAGCCAA 300
246 nLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeu 263
|||||
301 CAAGAAGGTGGCCAGCAATGCAGCCGAGCAACATGCTGAGAGATCCTTG 350
263 LyPheLysValProGlnAlaGlnProAlaLysProAlaLeuLysSergLup 279
|||||
351 GTTCAAGTCCCGCAGCGCAGCCACCAAAACCGCAGTCAAGTCAGAG 400
```

```
280 GluLysThrProValLysLysProGlyAspGlyArgLysValThrPhePh 296
|||||
|||||
401 GAGAAAGACACCCATAAGAAACAGGGGATGGAGAAAGTAACCTTTT 450
|||||
|||||
296 eGluProSerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluP 313
|||||
|||||
451 TGAACCTGGCTCTGGGGATGAACATGGGACTAGTACTAAAGAGATGAGT 500
|||||
|||||
313 heArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPro 329
|||||
|||||
501 TCAGGATGCCCTTATCTAGTCATCAAGCAGCTGCTGTTGGAATCTTCCC 550
|||||
|||||
330 MetValProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisTh 346
|||||
|||||
551 ATGGTGACCGACGCTGC.CAGGCTGTAGAGTCAGTCAAGACATCACAC 599
|||||
|||||
346 rLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThra 363
|||||
|||||
600 CACAGATT.ACAGGCCAGCATCCGAATCTGCCAAGACACACGGTACTG 648
|||||
|||||
363 IamettIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThraLaglu 379
|||||
|||||
649 CACAAGATAGCCCGAAGTGGGGAAGGGGCACCTCGACCA..... 689
|||||
|||||
380 ThrIleLeuLysSerAsnIleSerSerGlyHisValProHisGly..... 394
|||||
|||||
690 .....CAGCGAAACATATCAACAGACACACCTCTCTCAGGCAAGTA 730
|||||
|||||
395 .ProArg 396
|||||
731 ACCGAGG 737
```

